Quantitative Risk Assessment in Drill Casing Design for Oil and Gas Wells

Xutuan Zhang BSc, MSc

A thesis submitted in partial fulfilment of the
Requirements of the University of Wolverhampton
for the degree of Doctor of Philosophy

September 2004

Signature..... Xutuan Zhang .....

Date..... 19 - 11 - 2004 .....

Abstract

In the oil and gas industry the use of a reliability based design is becoming increasingly important because of the increasingly requirements for safety and economy. In contrast to the traditional working stress design (WSD), Quantitative Risk Assessment (QRA) provides the methodology for quantifying the risk of the design for a particular scenario.

The full QRA methodology was discussed and a mathematical model, based on Generalised Pareto Distribution (GPD) and Asymptotic important sampling (AIS) techniques, was built to give more precise answer by analysing limited random data points rather than using the assumed pre-defined distributions. Particular attention is paid to the tails of the distribution to obtain a good fit. The methods developed are compared with the traditional methods such as First/Second Order Reliability Method (FORM/SORM), Monte Carlo Simulation (MCS) to assess the efficiency and accuracy.

It is shown that, for the examples considered, the proposed methods provide accurate and efficient results for the probability of failure. Another important characteristic of this method is that it uses the random data and does not need the user to determine the distribution type of the variables. And the mathematical model built in the present research is a generalised method and can be use for other risk assessment.

Acknowledgement

Special thanks to ***Professor Musa Mihsei, Professor Richard Hall*** and ***Dr Kevin Kibble*** for their support, encouragement and knowledge that I have received from them all these years.

Thanks ***Dr Stewart Mcwilliam*** of University of Nottingham, and ***Dr Tahar Laoui*** of University of Wolverhampton for their review of the thesis and the suggestions they have been given.

Thanks to my wife and my parents, I wouldn't be able to finish my study without their support and understanding.

Table of Contents

Abbreviations

AIS	Asymptotic Important Sampling
CDF	Cumulative Distribution Function
COV	Coefficient of Variance
DCE	Design Check Equation
FORM	First Order Reliability Method
GPD	Generalised Pareto Distribution
ID	Inside Diameter
IS	Important Sampling
LS	Limit State
LSF	Limit State Function
LSS	Limit State Surface
MCS	Monte Carlo Simulation
ML	Maximum Likelihood
MRL	Mean Residual Life
OD	Outside Diameter
PDF	Probability Density Functions
PML	Point of Maximum Likelihood
PWM	Probability Weighted Moments
QRA	Quantitative Risk Assessment
SF	Safety Factor
SLS	Serviceability Limit State
SORM	Second Order Reliability Method

SWSE	Sum of Weighted Square Errors
TFP	Target Failure Probability
THI	Tail Heaviness Index
TRL	Tolerable Risk Level
ULS	Ultimate LS
WSD	Working Stress Design

List of Figures

List of Tables

Chapter 1 : Introduction

1.1 Introduction of the Oil and Gas Industry

The oil and gas industry plays one of the most important roles in the world. In recent years, the drilling of the oil and gas well has presented the industry with many problems. It is therefore particularly important for the study of drilling in petroleum engineering and development.

Drilling starts after the environmental and seismic surveys are complete, and drilling an exploratory well is considered if the results look good. Even at this stage it remains an uncertain business with no guarantees. There is still a high risk that nothing at all will be found or that the oil will be in such small quantities that it would not be worthwhile extracting. For example, in the North Sea only about one in eight exploration wells find quantities of oil and gas that are economic to develop. However, the aim of the present research is not about how to find the place to drill but how to drill it safely.

1.2 Oil and Gas Well Drilling

1.2.1 Drilling Rigs

There are three types of rigs which are used for different environment. In very deep sea, drilling ships are used (Figure 1-1(a)), for water depth between 360 metres and

100 metres, semi-submersible rigs are used (Figure 1-1(b)), and jack-up rigs used in shallow water less than 100 metres deep (Figure 1-1(c)).

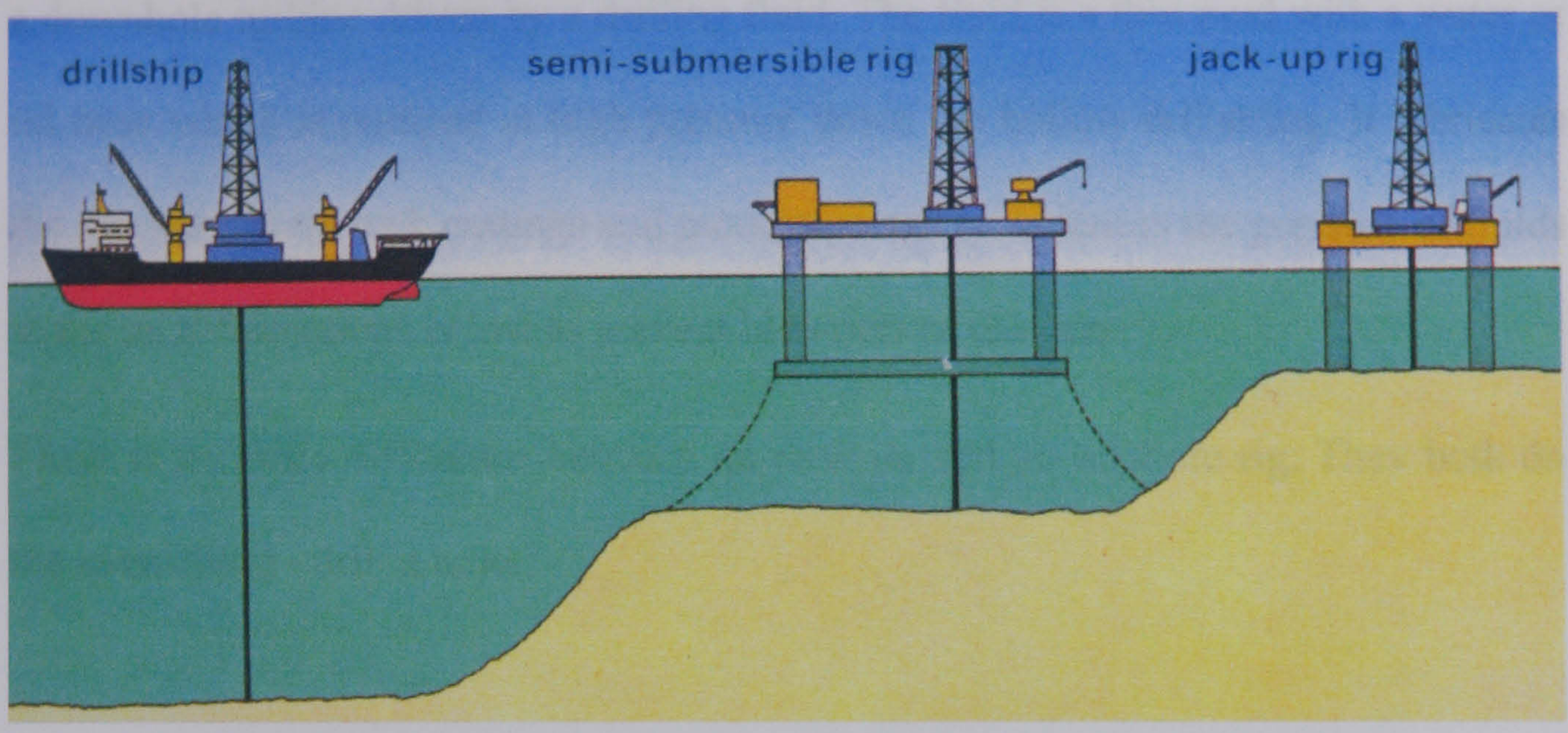


Figure 1-1: Drill Rigs: (a) Drilling Ships; (b) Semi-submersible rigs; (c) Jack-up rigs

Jack-up rigs have lattice legs that are lowered to the seabed before the floating section carrying the derrick is raised above the sea surface. Semi-submersible rigs float at all times, but when in position for drilling are anchored and ballasted to float lower in the water with their pontoons below wave-level. Some have dynamic-positioning propellers and can drill in very deep water.

The drilling rig itself is a derrick towering above the drill floor where most of the human activity is concentrated (Figure 1-2). The derrick supports the weight of the drillstring that is screwed together from nine metre lengths of drillpipe. Hoisting equipment in the derrick can raise or lower the drillstring up to three pipe lengths. At the bottom of the drillstring is a drill bit, which can vary in size and type. It is attached to the drill collars, the heavy pipe-sections that put weight on the bit. The rest of the drillstring is supported in tension by the derrick; otherwise, it would collapse under its own weight.

On semi-submersible rigs a compensator keeps the drillstring stationary while the rig and derrick move. The drill bit is rotated either by turning the whole drillstring or by a downhole turbine driven by a drilling fluid. The fluid is a thin mud with a water or oil base which is pumped at high pressure down the hollow drillstring. It lubricates the kit, washes up rock cuttings and most importantly, balances the pressure of fluids in the rock formations below to prevent blowouts or Gushers.

There is no basic difference between an onshore and an offshore rig. They both do the same thing - drill a hole.

1.2.2 Drilling Components

Figure 1-2 shows a typical drill tower, the following sections give a brief introduction to the main procedures and components.

a) Spudding the well

The first step is to drill a hole and then put a wide pipe into it to guide the drill and the drilling fluid. As each section of the well is completed it is lined with a heavy steel pipe casing that is cemented in place to prevent it caving in. This process is called spudding the well.

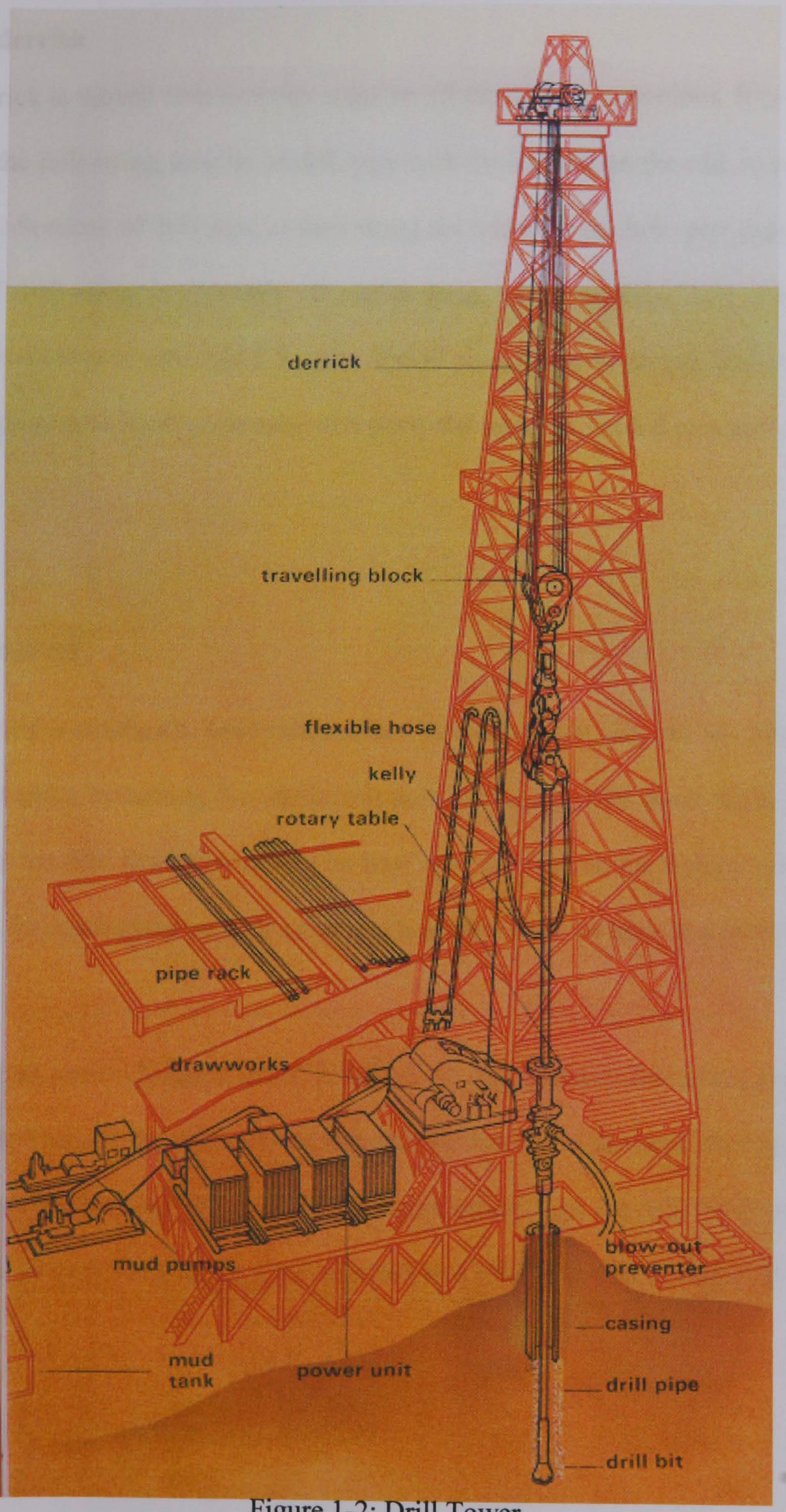


Figure 1-2: Drill Tower

b) The derrick

The derrick is the tall structure that supports all the drilling operations. It is designed to haul the drill string, lengths of drill pipe with the drill bit on the end, in and out of the hole. Sections of drill pipe or drill string are added as the hole gets deeper. Each piece of drill string is typically 10 metres long, which explains why the derrick, which needs to accommodate 3 lengths of drill pipe, has to be so tall. The drill string can weigh several hundred tonnes, so a powerful motor is needed to winch it up and down.

c) The drill bit

The drill bit is attached to heavy drill collars that put weight onto the bit. The drill bit is rotated either by turning the whole drill string or by a turbine down the hole that is driven by the drill fluid. Depending on how hard the rock is, the drilling rate can be less than 30 centimetres an hour or as much as 60 metres an hour in a relatively soft rock.

The drill bit needs changing every few days, or maybe every few hours, depending on the hardness of the rock. When the drill bit has to be changed, the whole drill string is pulled back up, uncoupled in sections, stacked up, the drill bit changed and the whole process starts again. This is known as a round trip and can take 10 hours or more.

d) Use of mud

Drilling fluid, which is commonly known as mud, is continuously pumped at high pressure down to the drill bit to lubricate it and keep it cool. The mud also flushes out the rock cuttings and brings them back to the surface. Geoscientists are able to inspect and analyse these tea leaf sized samples and gain more information about the rock structures and the presence of hydrocarbons. The person who does this is known as a mud-logger.

Another important consideration is that the force and weight of the mud that is pumped down the drill string into the well, balances the pressure of the crude oil and gas in the surrounding rocks and so significantly reduces the risk of a blow out.

e) Controlling the well

Because the oil and gas, deep below the Earth's surface are at high pressure, great care has to be taken to control the pressure. 'Blow out' can occur when a drill enters a reservoir and the pressure causes the oil and gas come spurting out of the well. The result is potentially very dangerous. Although blow outs are very unusual, all wells are fitted with an emergency valve designed to prevent this from happening.

f) Core samples

If the geologists find something particularly interesting, they can ask for a core sample. A hollow drill, called a core barrel, is attached to the drill string and as it goes down a core of rock forms inside. This core sample gives a continuous record of the different layers of rock and therefore more detailed information than the rock

cuttings. Collecting a core sample is expensive and time consuming because it involves a complete round trip.

g) Logging

Before a well is finally capped vital information is gathered by lowering measuring devices, contained a sonde, down the hole on a wire line. As the line is pulled back up the hole the sonde transmits information to a computer on the surface about the porosity and other qualities of the rock it is cutting through. This information provides a survey of the well and gives more information about the presence of hydrocarbons in the pores.

h) Field Appraisal

If an exploratory well shows that hydrocarbons are present, more seismic data is gathered and then a number of appraisal wells are drilled and more data is collected. From this data it is possible to estimate how much oil and gas the field contains, how difficult it will be to extract and what percentage of the oil and gas can be extracted.

1.2.3 Drilling

Once it has been established that a potential oil-bearing structure exists, the only way of confirming the existence of oil is to drill a well. In practice, the frequency of striking oil in an unexplored area is one in eight. In areas where there is a great deal of vegetation and soft ground, a stovepipe (762-1066.8mm OD) is driven by a pile driver to a depth of approximately 3048 mm. This is necessary to protect surface

formations from being eroded by the drilling mud, causing a large washout with eventual loss of the rig.

The two main tasks in drilling are:

(1) Adding fresh lengths (joints) of drill pipe as the drill bit bites down into the rock;

(2) The extraction of the entire drill string to change the bit or retrieve rock cores.

The oil well proper starts off with a hole size ranging from 444.5 to 914.4 mm and is drilled to a depth of 6,096-9,144 mm. The bottom hole assembly required to drill a large hole to a shallow depth normally consists of drill collars and one stabiliser. For deep holes a more rigid bottom hole assembly using three stabilisers is required to keep the hole straight or to maintain existing hole deviation. The basic bottom hole assembly consists of bit, near bit stabiliser, two drill collars, stabiliser, two or three drill collars, stabiliser, drill collars, HWDP and drill pipe to surface.

The first string of casing (339.725-728mm OD) is described as conductor pipe, and is run mainly to provide a conduit for the drilling mud.

Once the conductor pipe is cemented, a smaller bit, with a different bottom hole assembly, is run inside the conductor pipe and another hole (the surface hole) is drilled to a prescribed depth. The depth is dictated by hole conditions and formation pressures. Another casing string (the surface casing) is run and cemented.

The process of drilling a hole and running casing continues until the oil or gas zone is reached. The last casing string is described as a production string. Typical hole/casing sizes for a development area (i.e. where oil is known to exist from previous exploration drilling) are given in Table 1-1. It should be noted that combinations of hole/casing sizes other than those listed in Table 1-1 are also in use.

Table 1-1: Typical hole size/casing size arrangements

Hole Size (mm)	Casing Size (mm)	Description
914.4	762	Stove pipe
609.6/660.4	473.1/508	Conduct or pipe
444.5	339.7	Surface casing
311.2	244.5	Intermediate casing
215.9	177.8	Production casing
152.4	114.3/127	Full production casing or a production liner

1.2.4 Tripping operations

The term "tripping operations" refers to the process of lowering or raising the drill string into or out of the hole. The drill string is frequently pulled out of hole (POH) in order to change the drill bit or when the hole is drilled to its final Target Depth (TD) before the casing is run. Tripping in or running in hole (RIH) is the process of lowering the entire drill string after a bit change or for reaming and circulation purposes.

Figure 1-3 is a schematic drawing of the processes involved in tripping out. It starts by first raising the kelly above the rotary table, setting slips and then breaking the kelly, kelly bushing and swivel from the topmost joint and stacking them in the rathole, as shown in Figure 1-3.

The drill pipe is then removed by attaching pipe elevators to the drill pipe and using the drum hoist to raise the drill pipe above the rig floor. Basically, an elevator is a set of clamps that latches onto the pipe which allows the drill string to be lifted out of the hole.

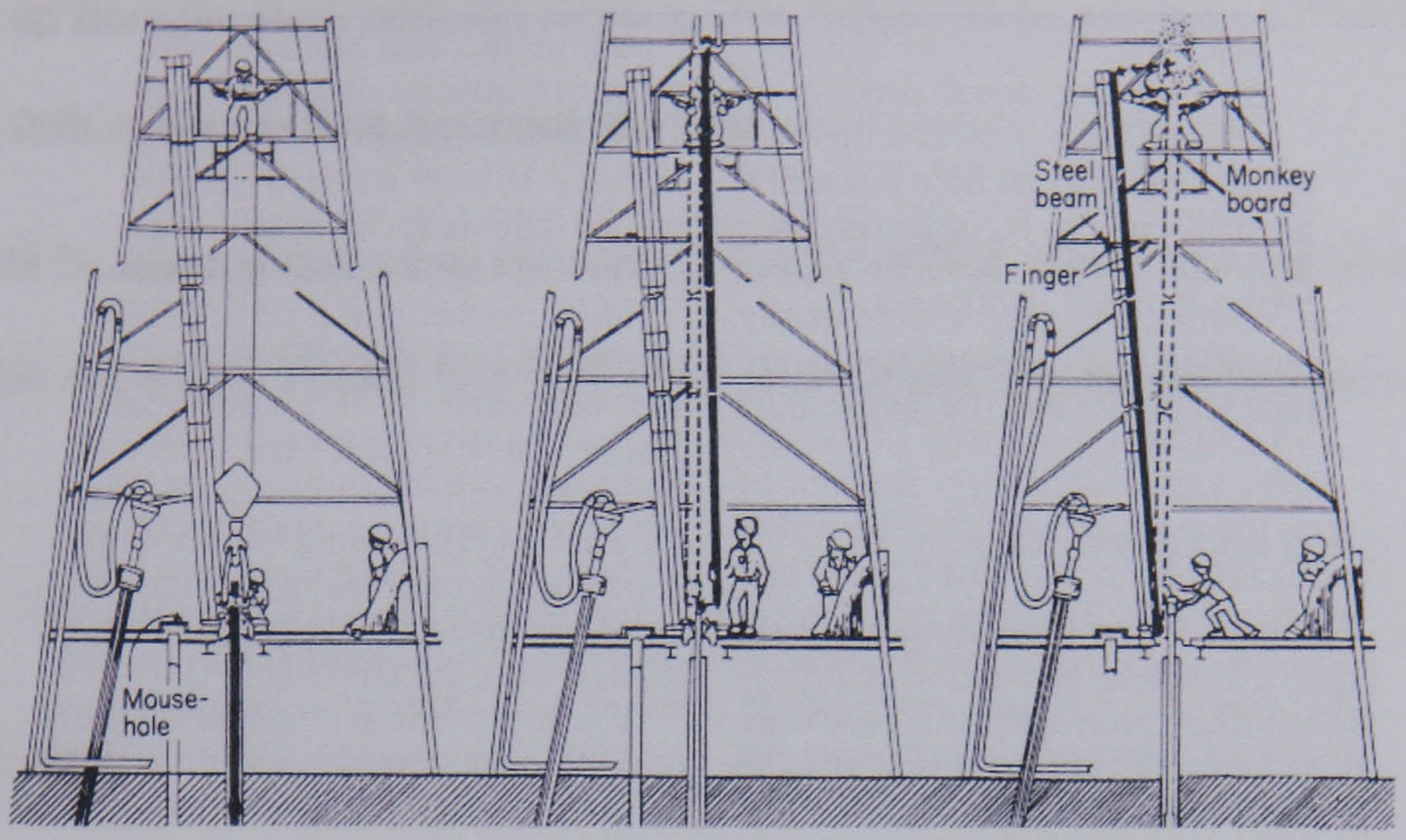


Figure 1-3: Tripping-out operation

The drill string is normally removed in a stack of three joints, described as a stand. A stand of drill pipe (approximately 2833.64 mm long) is raised above the rotary table and then disconnected, or broken, with tongs and spun out of hole with the pipe spinner or by back rotating the rotary table. The top of the stand is then picked up by a derrick man working on the platform (or monkey-board) from where he unlatches the pipe from the elevator. The top of the stand is then moved into a specially designed finger-board within the platform, just after the workmen (roughnecks) on the rig floor swing the lower part of the stand aside on the rig floor immediately below the monkey-board.

The empty elevators are then lowered and latched onto the top of the remaining drill string. The slips are removed from the rotary table and another stand is removed from the hole. This process is continued until the entire drill string is removed from the hole and stacked in the derrick.

Tripping-in operations use the same process as tripping out but in reverse, i.e. pipe is picked up from the stack with the elevator. The bottom hole assembly, including drill bit and drill collars, is first run inside the hole.

It should be noticed that when the well is finally drilled, tested and completed, the drill pipe stands are broken into individual joints inside the mouse-hole prior to rig move.

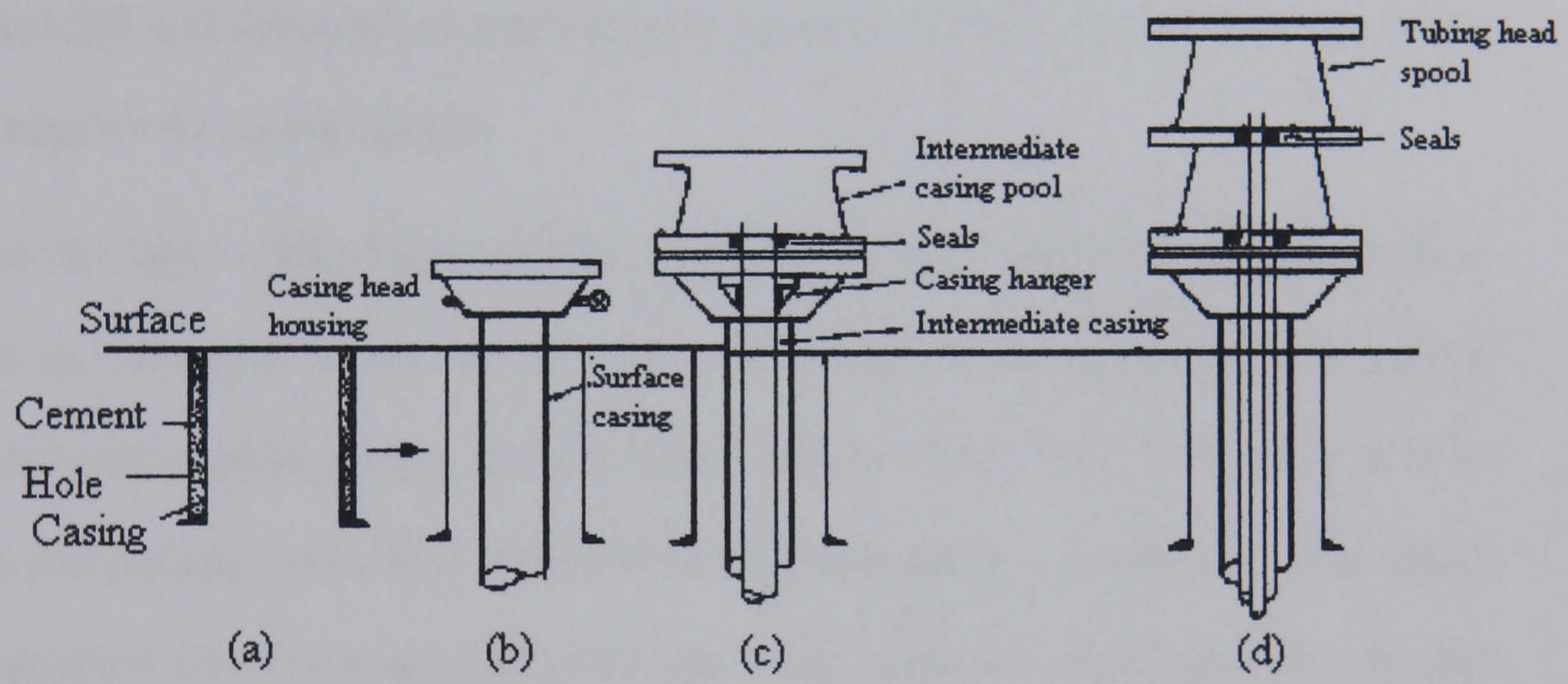
1.2.5 Casing

Casing is run and cemented after each section of the hole is drilled. Referring to Table 1-1, the conductor pipe is run and cemented to the surface. Normally this pipe does not carry any wellhead equipment (i.e. BOPs) and the drilling flowline is nippled directly to this string. (An exception to this is in offshore oil operations or areas where shallow gas may be encountered.) Drilling of the surface hole (e.g. 444.5mm) is carried out through the conductor pipe until the complete section is drilled. Surface casing (e.g. 339.725mm OD) is then run and cemented. In land drilling, the conductor pipe is cut so that it is level with the cellar floor.

The 339.725mm casing will carry the blowout preventers which are necessary to provide control while the next hole is being drilled. The 339.725mm casing includes a connection called casing head housing (CHH). Cementing the casing is facilitated by using a landing joint to which the cementing equipment is attached. Later the landing joint is removed and the CHH is screwed onto the topmost joint of the casing. In some cases a landing joint cannot be used, owing to difficulty in spacing out, and the cementing equipment will have to be attached to the casing itself. In this

case, after cementing, the casing is cut off and dressed. Then CHH is welded onto the top of the casing. The BOPs equipment can now be attached to the top of the CHH to allow safe drilling of the next hole. The casing head housing is also used to suspend the next string of casing. The wellhead arrangement after cementing the surface casing is shown in Figure 1-4.

Figure 1-4: Wellhead arrangement after setting: a) conductor pipe; b) surface casing;
c) intermediate casing; d) production casing.



The intermediate hole is drilled inside the surface casing. The intermediate casing (e.g. $9\frac{5}{8}$ in OD) is run and cemented in the same manner as the surface casing. This casing is then landed by hanging it inside the CHH using a special casing hanger. The casing hanger employs gripping teeth which hold the casing tightly in place; the casing hanger is also provided with teeth on the outside to engage with the CHH.

The top of the intermediate casing is then cut and dressed, and another spool, described as an "intermediate casing spool", is attached to the CHH. The intermediate spool provides the following functions:

(a) to pack off the top of the intermediate casing, thereby restricting well-bore fluid pressure to the inside of the casing;

(b) to carry the next stack of BOPs required to control hole during the drilling of the next hole;

(c) to provide suspension for the next casing string.

The CHH and intermediate spool(s) have recesses cut inside their bodies to provide seating for the casing hangers.

The last hole is drilled through the intermediate casing, and the producing string is run and cemented as before. The production string is hung inside the intermediate spool with a casing hanger. The production casing (114.3, 127 or 177.8 mm OD) is cut and dressed, and a final spool, the tubing head spool, is screwed on. The tubing head spool provides a pack-off to the top of the production string and also provides seating for the tubing hanger which will hold the production tubing. Oil or gas is normally produced through the tubing (rather than through the casing) which is sealed in a production packer placed just above the production zones.

1.2.6 Logging, testing and completion

After the well is drilled to TD, the hole is normally logged (in open or cased hole) by running specialised equipment on wireline. The primary object of open hole logging is to determine porosity, water saturation and boundaries of production zone(s).

These parameters are necessary in the determination of the quantity of movable oil and, in turn, the production life of the reservoir.

Completion of an oil well involves setting a production packer, running the production tubings and, finally, perforating the production zone(s). A production packer is set, just above the production zone, to seal the annulus from the reservoir pressure and confine fluids to the production tubing. The tubing is screwed at the surface to a tubing hanger; the latter is landed in the tubing head spool.

In areas where several oil reservoirs exist in the same well, dual production may be practised, where two tubing strings are run to the different production zones. Two packers are therefore necessary to seal both the production zones from the annulus.

Finally, a Christmas tree is attached to the top flange on the tubing head spool. The Christmas tree is a solid piece of steel with a hollow passage connected to the top of the tubing. It has a number of valves to control the flow of hydrocarbon from the well.

1.3 Casing Design

1.3.1 Purpose of Casing

As stated in previous sections, at a certain stage during the drilling of oil and gas wells it becomes necessary to line the walls of a borehole with steel pipe, known as the casing. Casing serves numerous purposes during the drilling and production history of oil and gas wells. These include:

1. Keeping the hole open by preventing the weak formations from collapsing, i.e. caving of the hole.

2. Serving as a high strength flow conduit to surface for both drilling and production fluids.

3. Protecting the freshwater-bearing formations from contamination by drilling and production fluids.

4. Providing a suitable support for wellhead equipment and blowout preventers for controlling subsurface pressure, and for the installation of tubing and subsurface equipment.

5. Providing safe passage for running wireline equipment.

6. Allowing isolated communication with the perforated formation(s) of interest.

1.3.2 Types of Casing

Hostile environments, such as high-pressured zones, weak and fractured formations, unconsolidated formations and sloughing shales, are often encountered. Consequently, wells are drilled and cased in several steps to seal off these troublesome zones and to allow drilling to the final depth. Different casing sizes are required for different depths, the five general casings used to complete a well are: conductor pipe, surface casing, intermediate casing, production casing and liner. As shown in Figure 1-5, these pipes are run to different depths and one or two of them may be omitted depending on the drilling conditions; they may also be run as liners or in combination with liners. In offshore platform operations, it is also necessary to run a cassion pipe.

- **Cassion Pipe**

A cassion pipe, usually 660.4 to 1066.8mm in outside diameter (OD), is driven into the seabed to prevent washouts of near-surface unconsolidated formations and to ensure the stability of the ground surface upon which the rig is seated. It also serves as a flow conduit for drilling fluid to the surface. The cassion pipe is tied back to the conductor or surface casing and usually does not carry any load.

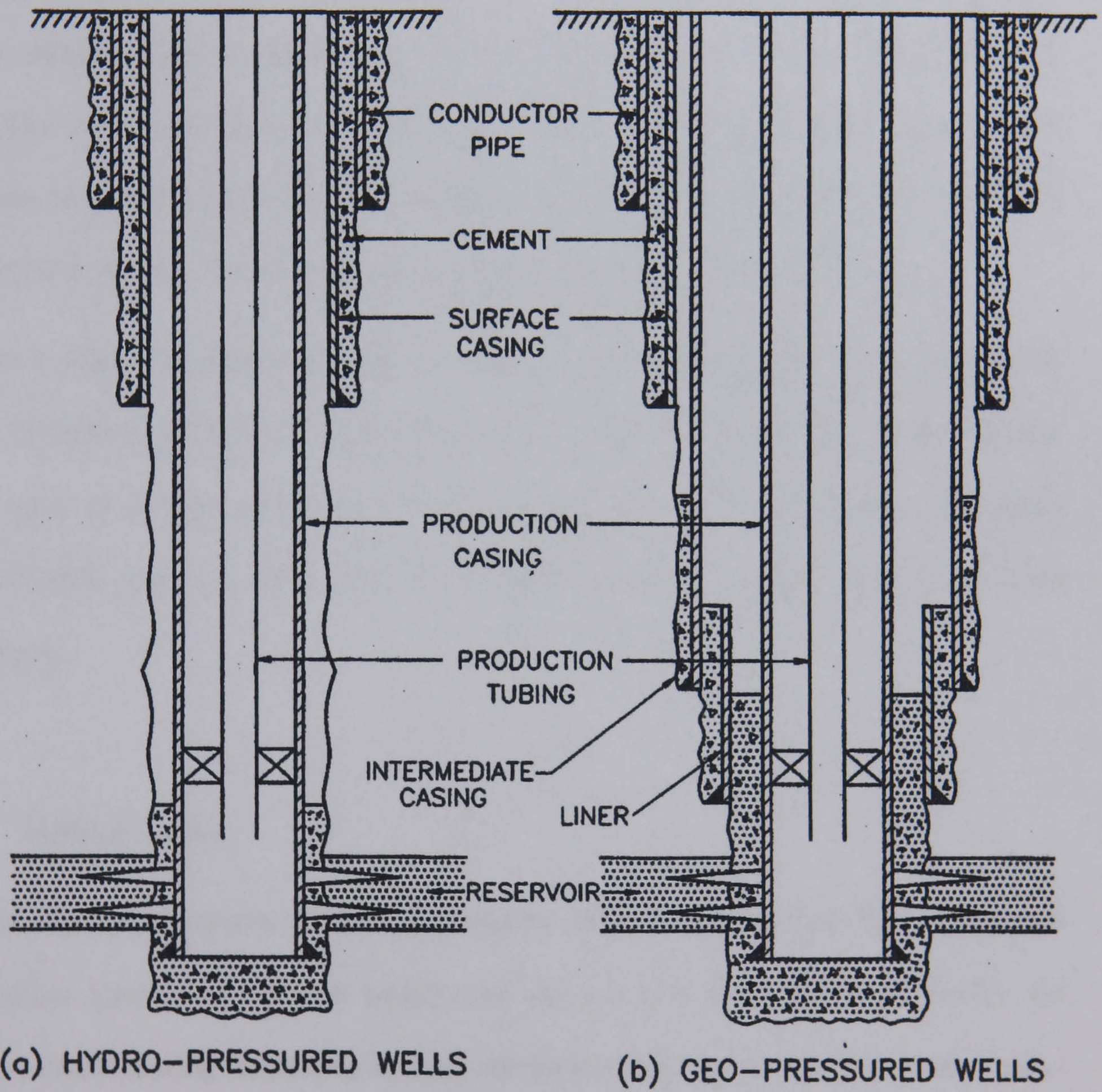


Figure 1-5: Typical casing program showing different casing sizes and their setting depths

- **Conductor Pipe**

The outermost casing string is the conductor pipe. The main purpose of this casing is to hold back the unconsolidated surface formations and prevent them from falling into the hole. The conductor pipe is cemented back to the surface and it is either used to support subsequent casings and wellhead equipment or the pipe is cut off at the surface after setting the surface casing. Where shallow water or gas flow is expected, the conductor pipe is fitted with a diverter system above the flow line outlet. This device permits the diversion of drilling fluid or gas flow away from the rig in the event of a surface blow-out. The conductor pipe is not shut-in in the event of fluid or gas flow, because it is not set deep enough to provide any holding force.

The conductor pipe, which varies in length from 1219.2 to 15240 cm onshore and up to 30,480 cm offshore, is 17.78 to 50.8 cm. in diameter. Generally, a 40.64 cm pipe is used in shallow wells and a 50.8 cm in deep wells. On offshore platforms conductor pipe is usually 50.8 cm in diameter and is cemented across its entire length.

- **Surface Casing**

The principal functions of the surface casing string are to: hold back unconsolidated shallow formations that can slough into the hole and cause problems, isolate the freshwater-bearing formations and prevent their contamination by fluids from deeper formations and to serve as a base on which to set the blow-out preventers. It is generally set in hard limestone or dolomite, so that it can hold any pressure that may be encountered between the surface casing seat and the next casing seat.

Setting depths of the surface casing vary from a few hundred feet to as much as 152,400 cm. The Size of the surface casing varies from 17.78 to 40.64 cm in diameter, with 27.305 cm and 33.9725 cm. being the most common sizes. Surface casing is usually cemented to the surface. For offshore wells, the cement column is frequently limited to the kick-off point.

- **Intermediate Casing**

Intermediate or protective casing is set at a depth between the surface and production casings. The main reason for setting intermediate casing is to case off the formations that prevent the well from being drilled to the total depth. Troublesome zones encountered include those with abnormal formation pressures, lost circulation, unstable shales and salt sections. When abnormal formation pressures are present in a deep section of the well, intermediate casing is set to protect formations below the surface casing from the pressures created by the drilling fluid specific weight (required to balance the abnormal pore pressure). Similarly, when normal pore pressures are found below sections having abnormal pore pressure, an additional intermediate casing may be set to allow for the use of more economical, lower specific weight, drilling fluids in the subsequent sections. After a troublesome lost circulation, unstable shale or salt section is penetrated, intermediate casing is required to prevent well problems while drilling below these sections.

Intermediate casing varies in length from 2,133.6 mitre to as much as 4,572 mitre and from 17.78 cm to 29.845 cm in outside diameter. It is commonly cemented up to 304.8 mitre from the casing shoe and hung onto the surface casing. Longer cement columns are sometimes necessary to prevent casing buckling.

- **Production Casing**

Production casing is set through the prospective productive zones except in the case of open-hole completions. It is usually designed to hold the maximal shut-in pressure of the producing formations and may be designed to withstand stimulating pressures during completion and workover operations. It also provides protection for the environment in the event of failure of the tubing string during production operations and allows for the production tubing to be repaired and replaced.

Production casing varies from 11.43 cm to 24.4475 cm in diameter, and is cemented far enough above the producing formations to provide additional support for subsurface equipment and to prevent casing buckling.

- **Liners**

Liners are the pipes that do not usually reach the surface, but are suspended from the bottom of the next largest casing string. Usually, they are set to seal off troublesome sections of the well or through the producing zones for economic reasons.

The major advantages of liners are that the reduced length and smaller diameter of the casing results in a more economical casing design than would otherwise be possible and they reduce the necessary suspending capacity of the drilling rig. However, possible leaks across the liner hanger and the difficulty in obtaining a good primary cement job due to the narrow annulus must be taken into consideration in a combination string with an intermediate casing and a liner.

1.3.3 Casing Properties

All specifications of casing include outside diameter, wall thickness, drift diameter, weight and steel grade. In recent years the API has developed standard specifications for casing, which have been accepted internationally by the petroleum industry.

Table 1-2: API manufacturing tolerance for casing outside diameter (After API Spec. 5CT. 1992)

Outside Diameter (mm)	Tolerance (mm)	
26.67~ 88.9	+2.38125	-0.79375
101.6~127	+2.778125	−0.75%OD
139.7~219.075	+3.175	−0.75%OD
≥ 244.475	+3.96875	−0.75%OD

Casing diameters and wall thickness are the most important properties that are used in casing design. As discussed previously, casing diameters range from 11.43 cm to 60.96 cm so they can be used in different sections (depths) of the well. The tolerances shown in Table 1-2, from API Spec. 5CT (1992), apply to the outside diameter (OD) of the casing immediately behind the upset for a distance of approximately 12.7 cm.

Casing manufacturers generally try to prevent the pipe from being undersized to ensure adequate thread run-out when machining a connection. As a result, most casing pipes are found to be within $\pm 0.75\%$ of the tolerance and are slightly oversized.

Inside diameter (ID) is specified in terms of wall thickness and drift diameter. The maximal inside diameter is, therefore, controlled by the combined tolerances for the outside diameter and the wall thickness. The minimal permissible pipe wall thickness is 87.5% of the nominal wall thickness, which in turn has a tolerance of -12.5 %.

The minimal inside diameter is controlled by the specified drift diameter. The drift diameter refers to the diameter of a cylindrical drift mandrel, Table 1-3, that can pass freely through the casing with a reasonable exerted force equivalent to the weight of the mandrel being used for the test (API Spec. .5CT, 1992). A bit of a size smaller than the drift diameter will pass through the pipe.

The difference between the inside diameter and the drift diameter can be explained by considering a 17.78 cm, 29.763 kg/m casing, with a wall thickness, t, of 6.9088 mm.

Inside diameter = OD - 2t

= 17.78-2 * 0.69088

= 16.39824 cm

Drift diameter = ID - 0.3175

= 16.39824 – 0.3175

= 16.08074 cm

Drift testing is usually carried out before the casing leaves the mill and immediately before running it into the well. Casing is tested throughout its entire length.

Table 1-3: API recommended dimensions for drift mandrels (After API Spec. 5CT. 1992)

Outside and liner (cm)	Length (cm)	Diameter (ID) (cm)
≤ 21.9075	15.24	ID- 0.3175
24.4475~33.99692	30.48	ID- 3.96875
≥ 40.64	30.48	ID- 4.7625

The lengths of pipe sections are specified by API RP 5B1 (1988), in three major ranges: R1, R2 and R3, as shown in Table 1-4.

Table 1-4: API standard lengths of casing (After API RP 5B1, 1988)

Range	Length (cm)	Average length (cm)
R1	487.68~762	670.56
R2	762~1036.32	944.88
R3	Over 1036.32	1280.16

Generally, casing is run in R3 lengths to reduce the number of connections in the string, a factor that minimises both rig time and the likelihood of joint failure in the string during the life of the well. R3 is also easy to handle on most rigs because it has a single joint.

The steel grade of the casing relates to the tensile strength of the steel from which the casing is made. The steel grade is expressed as a code number which consists of a letter and a number, such as N-80. The letter is arbitrarily selected to provide a unique designation for each grade of casing. The number designates the minimal

yield strength of the steel in thousands of psi. Strengths of API steel grades are given in Table 1-5.

Table 1-5: Strengths of API steel grades. (API Spec. 5CT, 1992.)

API Grade	Yield Strength (Pa)		Minimum Ultimate Tensile Strength (Pa)	Minimum Elongation (%)
	Minimum	Maximum		
H-40	275,790,400	551,580,800	413,685,600	29.5
J-55	379,211,800	551,580,800	517,107,000	24.0
K-55	379,211,800	551,580,800	655,002,200	19.5
L-80	551,580,800	655,002,200	655,002,200	19.5
N-80	551,580,800	758,423,600	689,476,000	18.5
C-90	620,528,400	723,949,800	689,476,000	18.5
C-95	655,002,200	758,423,600	723,949,800	18.0
T-95	655,002,200	758,423,600	723,949,800	18.0
P-110	758,423,600	965,266,400	861,845,000	15.0
Q-125	861,845,000	1,034,214,000	930,792,600	14.0

* Elongation in 50.8mm, minimum per cent for a test specimen with an area ≥19.05mm.

Hardness of the steel pipe is a critical property especially when used in H2S (sour) environments. The L-grade pipe has the same yield strength as the N-grade, but the N-grade pipe may exceed 22 Rockwell hardness and is, therefore, not suitable for H2S service. For sour service, the L-grade pipe with a hardness of 22 or less, or the C-grade pipe can be used.

Table 1-6: Strengths of non-API steel grades

Non-API Grade	Manufacturer	Yield Strength ($\times 10^8$ Pa) Minimum	Yield Strength ($\times 10^8$ Pa) Maximum	Minimum Ultimate Tensile Strength ($\times 10^8$ Pa)	Minimum Elongation (%)
S-80	Lone Star Steel	5.17107**	-	5.17107	20.0
		3.79212***	-		
Mod. N-80	Mannesmann	5.51581	6.55002	6.89476	24.0
C-90◆	Mannesmann	6.20528	7.23950	8.27371	26.0
SS-95	Lone Star Steel	6.55002**	-	6.55002	18.0
		5.17107***	-		
S00—95		6.55002	7.58424	7.58424	20.0
S-95	Lone Star Steel	6.55002**	-	7.58424	16.0
		6.34318***	-		
S00-125	Mannesmann	8.61845	10.34214	9.30793	18.0
S00-140	Mannesmann	9.65266	11.37635	10.34214	17.0
V-150	U.S. Steel	10.34214	12.41057	11.03162	14.0
S00-155	Mannesmann	10.68688	12.41057	11.37635	12.0

* Test specimen with area greater than 483.87 mm^2.

** Circumferential

*** Longitudinal

◆ Maximal ultimate tensile strength of 8.2737 $\times 10^8$ Pa.

Many non-API grades of pipes are available and widely used in the drilling industry. The strengths of some commonly used non-API grades are presented in Table 1-6.

These steel grades are used for special applications that require very high tensile strength, special collapse resistance or other properties that make steel more resistant to H_2S.

Appendix D lists some most frequently used casing properties. More detail can be found in "Fundamentals of Casing Design"(Hussain Rabia, 1987).

1.3.4 Principles of Traditional Casing Design

The design of a casing program involves the selection of setting depths, casing sizes and grades of steel that will allow for the safe drilling and completion of a well to the desired production configuration. Very often the selection of these design parameters is controlled by a number of factors, such as geological conditions, hole problems, number and sizes of production tubing, types of artificial lift equipment that may eventually be placed in the well, company policy, and in many cases, government regulations.

Of the many approaches to casing design that have been developed over the years, most are based on the concept of maximum load. In this method, a casing string is designed to withstand the parting of casing, burst, collapse, corrosion and other problems associated with the drilling conditions. To obtain the most economical design, casing strings often consist of multiple sections of different steel grades, wall thickness, and coupling types. Such a casing string is called a combination string.

1.3.4.1 Factors influencing Casing Design

Casing design involves the determination of factors which influence the failure of casing and the selection of the most suitable casing grades and weights for a specific operation, both safely and economically. The casing program should also reflect the completion and production requirements.

A sound knowledge of stress analysis and the ability to apply it are necessary for the design of casing strings. The end product of such a design is a "pressure vessel" capable of withstanding the expected internal and external pressures and axial loading. Hole irregularities further subject the casing to bending forces which must be considered during the selection of casing grades.

A safety margin is always included in casing design, to allow for future deterioration of the casing and for other unknown forces which may be encountered, including corrosion, wear and thermal effects.

Casing design is also influenced by:

(a) Loading conditions during drilling and production;

(b) The strength of the casing seat (i.e. formation strength at casing shoe);

(c) The degree of deterioration which the pipe will be allowed during the entire life of the well;

(d) The availability of casing.

In general, the cost of a given casing grade is proportional to its weight, the heaviest weight being the most expensive. Since the cost of casing in a given well constitutes a high percentage of the total cost of drilling and completion, the designer should

ensure that the lowest grades and lightest weights, consistent with safety, are chosen as these provide the cheapest casing.

Incorrect design of the casing can result in disastrous consequences placing human lives at risk and causing damage and loss of expensive equipment. The entire oil reservoir may be placed at risk if the casing cannot contain a kick wasting a great deal of the reservoir's natural energy.

1.3.4.2 Traditional Design Criteria

The criteria for casing design are as follows:

a) Axial Tension

Most axial tension arises from the weight of the casing itself. Other tension loading can arise due to bending, drag, shock loading and during pressure testing of casing. In casing design, the uppermost joint of the string is considered the weakest in tension, as it has to carry the total weight of the casing string. Selection is normally based on a safety factor of 1.6 to 1.8 for the top joint (Hussain Rabia, 1987).

b) Collapse Pressure

Collapse pressure originates from the column of mud used to drill the hole, and acts on the outside of the casing. Since the hydrostatic pressure of a column of mud increases with depth, collapse pressure is highest at the bottom and zero at the top. Therefore:

Collapse pressure (C)=mud density × depth × acceleration due to gravity

$$= \rho \times g \times h$$

$$C = \rho h / 144 \tag{1-1}$$

where ρ is in lbm/ft^3, and h is in ft, or

$$C = 0.052 \rho h \text{ psi} \tag{1-2}$$

where ρ is in ppg.

In SI units, collapse pressure (C) is given by:

$$C = 9.8 \rho h \times 10^6 \text{ Pa} \tag{1-3}$$

where ρ is in kg/m^3, and h is in m.

The designer should ensure that the collapse pressure (C) never exceeds the collapse resistance of the casing. For this purpose, the casing collapse resistance is taken as the load at which the internal diameter of casing yields. In designing for collapse, the casing is assumed empty for surface and production casing and partially empty for intermediate casing.

For directional wells a correct well profile is required to determine the true vertical depth. Collapse pressure should be calculated using true vertical depth only.

c) Burst Pressure

The burst criterion in casing design is normally based on the maximum formation pressure resulting from a kick during the drilling of the next hole section. For added safety it is also assumed that the influx fluid(s) will displace the entire drilling mud,

thereby subjecting the inside of the casing to the bursting effects of formation pressure.

At the top of the hole the external pressure due to the hydrostatic head of mud is zero and the internal pressure must be supported entirely by the casing body. Therefore, burst pressure is highest at the top and lowest at the casing shoe where internal pressures are resisted by the external pressure originating from fluids outside the casing. As will be shown later, in production casing the burst pressure at shoe can be higher than the burst pressure at surface in situations where the production tubing leaks gas into the casing.

d) Compression Load

Compression loading arises in casings that carry inner strings where the weight of inner strings is transferred to the larger supporting casing. Production casings do not carry inner casing strings and are used, amongst other things, to suspend production tubings which are very light in comparison with casing. Consequently, production casings are not designed for compression loading.

e) Other Loadings

Other loadings that may develop in the casing include:

(a) bending with tongs during make-up;

(b) pull-out of the joint and slip crushing;

(c) corrosion and fatigue failure, both of the body and of the threads;

(d) pipe wear due to running wire line tools and drillstring assembly which can be extremely detrimental to casing in deviated and dog-legged holes; and

(e) additional loadings arising from treatment operations. The latter operations include acidising, squeeze cementing and hydro-fracturing.

1.3.4.3 Safety Factors

It is evident from the discussion of casing design criteria that exact values of loadings are difficult to determine. For example, if mud of 11,310 Pa/m is assumed to exist on the outside of the casing during running of the casing, this value cannot be expected to remain constant throughout the life of the well. Deterioration of mud with time will reduce this value to, say, a saltwater gradient of 10,528 Pa/m. Hence, calculations of burst pressures assuming a 11,310 Pa/m column of mud on the outside of the casing are not applicable throughout the life of the well. If the initial design of the casing is marginal, then any change in loading conditions may lead to bursting of the casing in the event of a gas leak from tubing during production operations.

Therefore, casing design is not an exact technique, because of the uncertainties in determining the actual loadings and also because of the change in casing properties with time, resulting from corrosion and wear. A safety factor is used to allow for such uncertainties and to ensure that the rated performance of the casing is always greater than any expected loading. In other words, the casing strength is down-rated by a chosen safety factor.

Each operating company uses its own values of safety factors for specific situations. These have been developed through many years of drilling and production experience. According to DEA(E) 64, working stress design uses safety factors at the following range:

Collapse: 0.85-1.125

Burst: 1 ~ 1.1

Tension: 1.6 ~ 1.8

The safety factor is determined as the ratio between body resistance and the magnitude of applied pressure. Thus safety factor (SF) in burst is given by:

$$SF = \frac{\text{burst resistance of casing}}{\text{burst pressure}}$$

In most circumstance, casing is designed to produce specified minimum safety factors for the three types of loading mentioned above. The casing grades will therefore have these minimum safety factors at the critical points along the hole. At other points the casing will have higher values for the safety factors.

1.4 Discussion

1.4.1 Limitations of Traditional Design Criteria

In casing design, safety and cost are two of the most important factors. Throughout industry reliability-based design methods have a long pedigree in control and optimisation of these factors.

Traditional casing design uses the deterministic approach, Working Stress Design (WSD). It consists of matching the nominal strength of the pipe with an estimated worst case accidental load, that is to say, in this kind of design minimum expected strength is designed to exceed maximum possible loads. Using this procedure the uncertainties in the design variables are accounted for by the use of design factors

(DFs). The origin of such factors is almost entirely empirical; these design factors have been handed down through the years; they have rarely been subjected to field or analytical verification. Also, DFs and loading considerations vary greatly within the industry. As a result, the true safety margin in various applications of the design is largely unknown.

In summary, this approach has a number of shortcomings, such as

1) Poor Economics: The present DFs are based on historical experience, but this is always of the 'worst' wells (e.g., wildcats), for which the failure probability is highest. DFs, which are correct for the worst cases, may be over-conservative for less severe cases such as development wells, resulting in wasteful over-sizing.

2) Inflexibility: In general, one DF is used for all strings and all load types, e.g., 1.2 is used for production casing burst after tubing leak as well as for pressure test of the outer casings. However, the consequences of the first scenario (possible release of hydrocarbons to the environment) are much more severe than the consequences of the second (casing repair); so the first condition should have a higher DF than the second.

3) Uneven risk: At present, the same DF is normally used for all steel types and grades. This is open to criticism as the margin of safety depends in part on the difference between ultimate tensile strength (UTS) and yield; and this is much lower for high strength steels (eg, V150) than for lower grades (e.g., J55, L80). Furthermore, W.S. Whitney (May 1995) and K.C. Maes (1993) that historical experience is almost solely for API steels, and hence the DFs based upon it may not be applicable to newer steel types.

1.4.2 Summary

In 1993, British Gas decided in DEA(E)-64 that it would be advantageous to explore the possibility of developing a new casing design methodology, based on probabilistic concepts and historical design data, and as is the case with all the design criteria used for mechanical and civil engineering systems (e.g. steel structures, concrete structures, foundation connections, vessels, pipelines, etc.), the objective of design is not to eliminate failures or to achieve absolute safety; rather, it is to strike an optimum balance between cost and risk; between economy, and safety and reliability.

In order to achieve a proper balance, it is necessary to focus on several important issues:

- What are the consequences of casing failures? For instance, it is clear that a lost return causes less human and environmental risk than a kick. A small economic losses must be considered.
- Which types of casing failure can be distinguished? Besides strength criteria, do certain serviceability criteria need to be taken into account?
- What is the probability of occurrence of a certain type of load, such as a kick load or a lost return?
- Which variables govern the behaviour and the response of casing systems?
- Are suitable mechanical/metallurgical models and reliable data available to calibrate design check equations? The consensus is that many current design criteria are too conservative, that they should rely on appropriate experimental results and sound mechanical models.

- How much risk is the designer willing to take?

As a fast growing statistical method, Quantitative Risk Assessment (QRA) methods provides the methodology for the optimisation of a design, while quantifying the risk of the design for a particular scenario and can give the designer a satisfactory and quantified answer.

The present research is to apply the full QRA methodology to casing design and to obtain more accurate results based on the limited random data set rather than using assumed pre-defined distributions for the variables involved. Traditional casing design methodology and its simulation methods such as First/Second Order Reliability Method (FORM/SORM), Monte Carlo Simulation (MCS) will be discussed in the following sections. Generalised Pareto Distribution (GPD) and important sampling (IS) techniques will be used to develop the new method.

Chapter 2 : Quantitative Risk Assessment

2.1 Introduction to Quantitative Risk Assessment (QRA)

Risk and uncertainty are essential features of most reliability problems and need to be understood for rational decisions to be made. Quantitative risk analysis offers the user a powerful and precise method for assimilating the various uncertainties of a problem and producing a realistic appreciation of the problem's total uncertainty.

The use of QRA technology is applicable for strength issues (worn casing, poor mechanical properties, poor dimensional properties, etc.) or load issues (unusual loading requirements during stimulation, fracturing, compaction, leaks, etc.). The risk associated with a particular load strength scenario can be quantified and balanced against the associated cost of the repair or operational change. The quantification of risk and the cost associated with that risk is paramount in the decision making process. Without the risk quantified, the engineer cannot properly evaluate the situation and come to the optimal risk-cost balanced decision.

Reliability-based analysis incorporates the variability on the strength side of the limit state function as well as the variability on the load side of the function. Once the load variabilities are defined in terms of a distribution (based on historical data), sophisticated reliability based simulations (due to the extremely low probabilities of failures which often occur) are performed which combines all the possible combinations of load and strength variables through the use of a Limit State Function (LSF).

In the QRA process, uncertainties that naturally occur in the determination of loads and casing strengths are explicitly taken into account. QRA allows for uncertainties to be considered in the analysis of specific wells. For each type of loading on a well, QRA takes into account the uncertainty of the particular loading based on actual field load data and the probability of a load event occurring. A probabilistic representation of each random variable describes the uncertainties including probability of occurrence, unavoidable scatter, and subjective modelling uncertainties. Uncertainties are measured by the statistical spread in the data.

QRA considers three major components: uncertainties, risk, and economics. Risk expresses the probability of an unfavourable outcome. The reliability analysis model invariably defines both load and strengths as probabilistic random variables. Risk depends on the degree of overlap of the load and Strength probability density curves. It is this overlap that defines the balance between load and design and therefore risk. A sophisticated evaluation approach is necessary to fully define the extreme values in the "tails" of both load and resistance distributions to ensure extreme values are implicit in the design premise. An important point is that there is no risk free environment.

Economics must enter the decision process since there is no zero risk operation. Higher safety margins will move apart and reduce (but not eliminate) the load and strength overlap. Evaluation of the relationship of cost versus risk in principle provides a direct means to assess problem situations. The limitation in applying this direct approach is that there are insufficient available data with which to model the distributions of loads and design. A flow chart shown in Figure 2-1 details the components and interaction in the analysis.

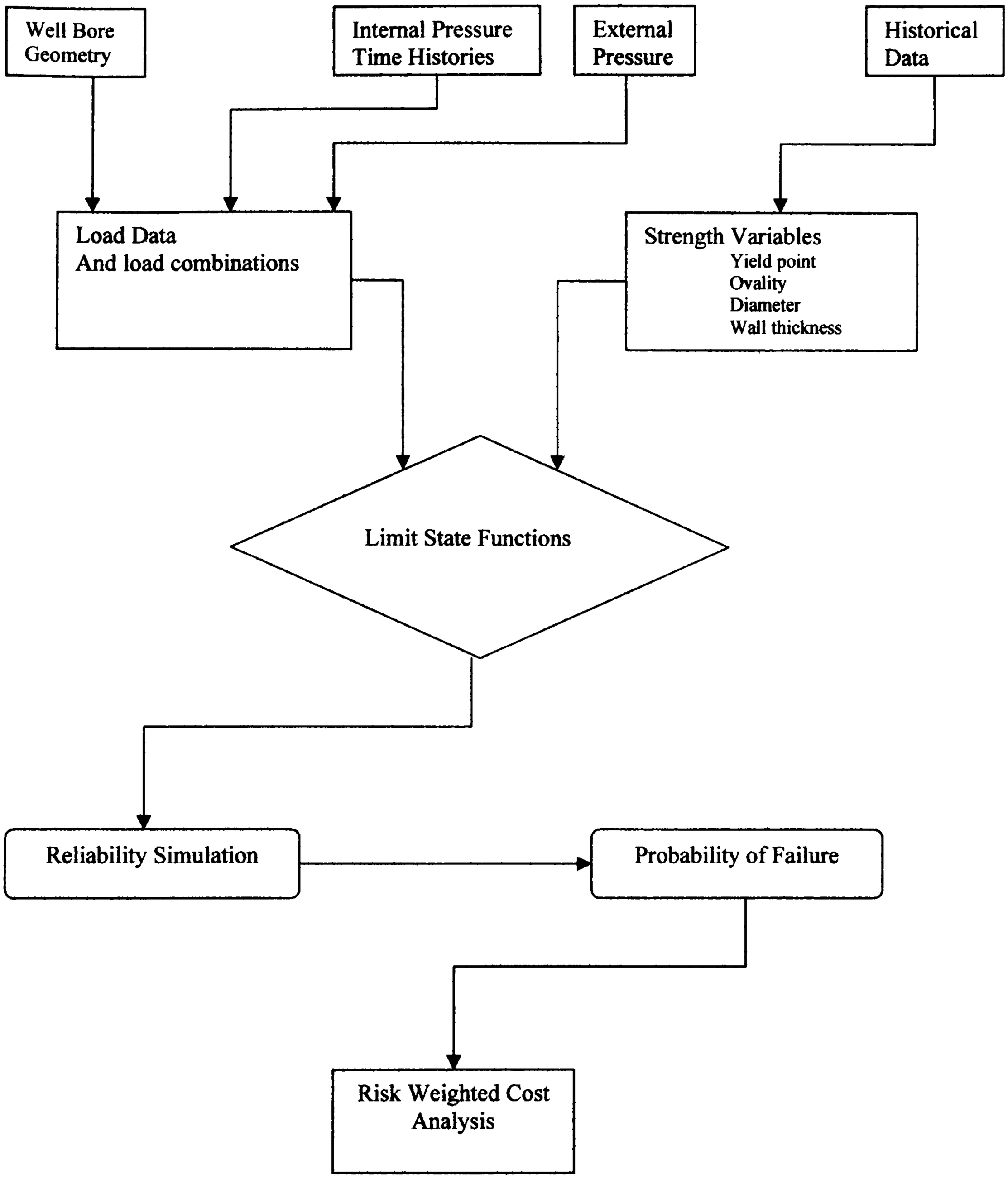


Figure 2-1: QRA flow chart

2.2 QRA in Casing Design

2.2.1 The advantages of using QRA in casing design

Well casing experiences operating loads due to running, cementing and testing operations and accidental loads due to formation/reservoir fluids upsurge in the wellbore (kick) or lost circulation caused by formation fracture. Large temperature variations, corrosive environments and pipe wear, especially at intended and unintended deviations from straightness are often encountered. The dominant casing failure modes consist of ductile/brittle under internal casing pressure, pipe instability or collapse under external pressure, tension or compressive/buckling failures of the casing string, and connection failures.

As is the case for all the design criteria in mechanical and civil engineering systems (e.g. steel structures, concrete structures, foundations, connections, vessels, pipelines, etc.), the objective of design is not to eliminate failures or to achieve absolute safety; rather, it is to strike an optimum balance between cost and risk; between economy and safety and reliability.

The ability of a risk management based design philosophy to reduce costs is widely recognised by industry (Gulati, D.L. McKenna, M.A. Maes, May 1994). Recently, the need for cost reduction in casing design due to competitiveness in today's market has prompted a closer investigation of the application of QRA principles in connection with casing design criteria. For example, tubular goods are estimated to represent an annual domestic rig cost of about $3.5 billion, which is 16% of the total

annual domestic drilling expenditure (Gulati, D.L. McKenna, M.A. Maes, May 1994).

Because of the magnitude of this cost, it is clear that small changes in the design criteria (e.g., safety factors, and choice of design load) have a huge impact on the economics of drilling. The current design framework cannot be used as a basis for improving the consistency of design criteria; QRA is needed as a fundamental principle in developing a reliability-based design format.

A first attempt to actually apply QRA to casing design system is described by T.B. Reeves et al. (1993). Well-established structural reliability techniques are used to calculate design factors for exploration and development wells. Even though the basic principles of LSD are not strictly adhered to, the analysis method proves to be extremely successful, and it also confirms the fact that the current design factors, which have evolved on a historical basis, constitute a worst design scenario for wells and/or load cases.

The advantages of using QRA in casing design is that it can address the vital issue of the consequences of failure, and how this should be included in the design philosophy. It gives the designer an answer to the 'how much' question and allows designer to calculate the probability of failure for any given structure and load case: or conversely, to obtain the design factors required to keep the failure probability below some given level.

It therefore appears as if QRA is justified by the benefits it offers in term of a better design. Furthermore, as the QRA process is used to calibrate the existing deterministic methods, it does not require any additional analysis at user level. The

user therefore achieves the twin benefits of simple analysis methods and better, more economic design.

In casing design, a practical reliability-based design format is based on Limit States Design (LSD). In summary, it has the following characteristics (K.C. Maes, P.R. Brand et al.):

- it gives engineers a better understanding of fundamental design principles and it enables them to exercise better judgement for designs using new materials and unusual operating condition;
- it provides a consistent approach for evaluating specific designs;
- it results in important cost savings, failure risks being more consistent with the consequences of failure;
- it provides society with a method by which human and environmental safety can be evaluated.

2.2.2 The Basic Approach: Limit States Design

Assume that casing systems and components are considered to be unfit for use (to have failed) when they exceed a particular state, called a Limit State (LS), beyond which their performance, or use is impaired. There are essentially two kinds of LS:

- Ultimate LS(ULS), which correspond with maximum load-carrying capacity and are considered fatal to the system; and

- Serviceability LS (SLS) which restrict normal use or affect durability.

An example of the former LS is burst in or buckling, whereas excessive deformation or leaking may be considered to be a SLS. For each LS a mathematical model needs to be established which incorporates the appropriate variables describing uncertainties with respect to loads, structural response, geometry, the interaction of the casing components, resistance, workmanship, environment, etc. These basic variables are the random variables $Z=(Z_1, ..., Z_n)$. For given design parameters such as nominal wall thickness or nominal mud weight, the variables Z can be modelled using probability density functions (PDFs). The LS function G can generally be formulated in terms of a load effect contribution L and a resistance contribution R:

$$G(Z) = R(Z) - L(Z) \tag{2-1}$$

and it must be chosen in such a way that negative values of G indicate that the combination $(Z_1, Z_2, ..., Z_n)$ results in a failure, that is, an exceedance of the LS:

$$G(Z)\begin{cases} > 0 : Z \text{ belongs to the safe set} \\ = 0 : Z \text{ lies on the LS surface} \\ < 0 : Z \text{ belongs to the failureset} \end{cases} \tag{2-2}$$

The probability of failure P(F) for a particular LS is consequently equal to

$$P(F) = P(G(Z) \le 0) = \int_{G(Z)\le 0} f(z)dz \tag{2-3}$$

where f(z) is the joint PDF of all the basic variables.

Also, the designer must be provided with relatively straightforward deterministic equations in order to check whether a particular LS is safe, this is achieved using a set of Design Check Equations (DCE) which are generally of the type:

Factored resistance r>Factored load effect q

but in a more general context, they can be expressed as:

$$\text{DCE: } r(z^*, \alpha) \geq q(z^*, \alpha) \tag{2-4}$$

where $z^* = (z_1^*, z_2^*, \cdots, z_n^*)$ is the set of (specified) design values (or, nominal values) associated with each of the basic variables $Z_i (i = 1, \cdots, n)$ and $\alpha = \{\alpha_1, \alpha_2, \cdots, \alpha_m\}$ is an array of Partial Factors (PFs). The role of the partial factors is to ensure that when a DCE is used for checking a particular LS, this results in a design having a failure probability (2-3), which is roughly equal to a predetermined value p^*. This small failure probability, p^*, is called the Target Failure Probability (TFP). It is a function of the acceptable risk level, the consequence of failure, and the interaction of several LS's.

2.2.3 Randomly Distributed Variables

Both the load and the strength capacity of a structural member are random in nature. In exploration well design, for example, an important loading conditions arises when the internal drilling fluids are accidentally replaced by a fluid from a high pressure zone below the casing shown. The load corresponding to this accidental condition, known as kick, is estimated by considering the drilling fluid as a particularly light

gas, methane. Actual measurements of kick loads show that both the magnitude and the occurrence frequency of kick are at considerable variance with the assumed design loads. Similarly, the estimated strength capacity of casing is normally different than the actual capacity due to uncertainties in material properties and dimensional tolerances.

Therefore it can be seen that most variables, in reality, do not have one single absolute value, as used in the deterministic approach. Take the example of yield strength for the casing. If a tensile test was performed on a single specimen, then a single value of yield strength would be obtained. However, if a number of these tests were repeated for an ostensibly identical specimen, then a range of values would be expected, as shown in Figure 2-2.

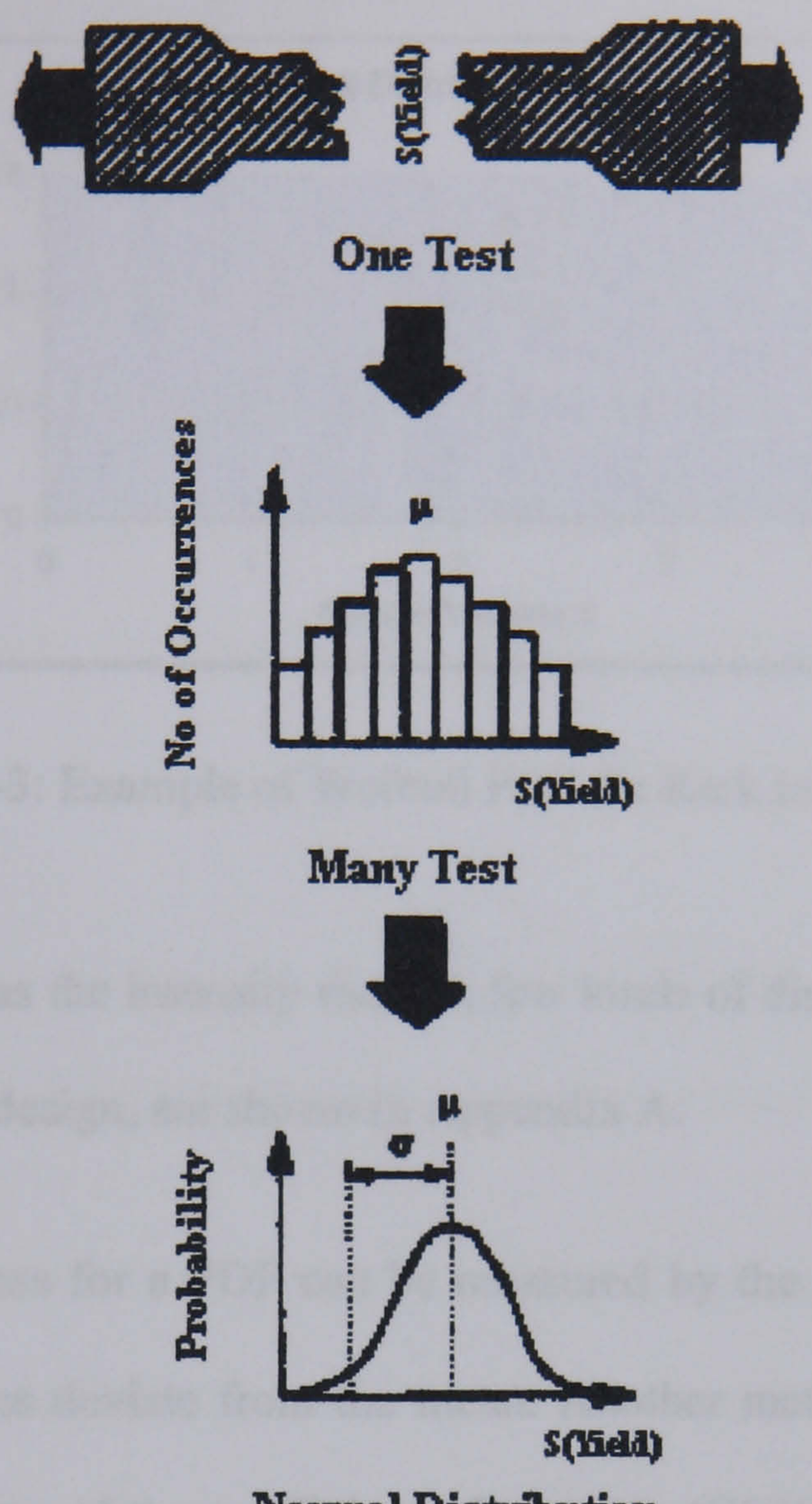


Figure 2-2: Example of a PDF

If the yield stress is plotted against the number of occurrences, then a bell-shaped curve, or probability distribution function (PDF) would be formed, showing a peak at the mean value of yield stress, μ. This bell-shaped curve is a good example of a normal distribution. And characterises many engineering quantities, but it is not the only type of PDF. For example, the Weibull distribution, shown in Figure 2-3, often describes gas kick intensity, with more kicks of lower intensity, and the frequency of

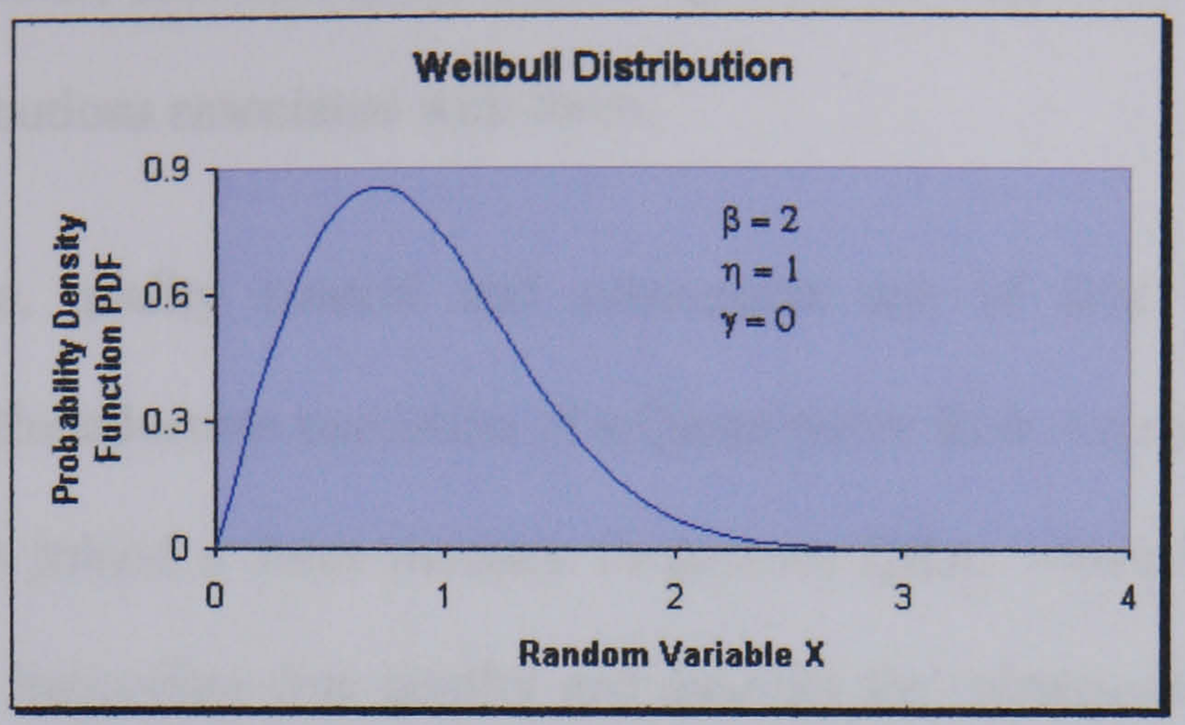


Figure 2-3: Example of Weibull PDF for Kick Intensity

occurrence tailing off as the intensity rises. A few kinds of distribution types, which often appear in casing design, are shown in Appendix A.

The spread of the values for a PDF can be measured by the standard deviation, σ, showing how far values deviate from the mean. Another method of expressing the spread is through the use of the coefficient of variance (COV), which combines the mean and standard deviations into a single dimensionless coefficient, often useful where only limited or generic data is available, or as a comparative indicator of dispersion.

$$\mathrm{COV} = \frac{\sigma}{\mu} \tag{2-5}$$

Each of the input variables in a casing design has a PDF associated with it. For example, on the load side, there is a pore pressure (predicted against actual) distribution, a kick size and intensity distribution. Also, when predicting casing

resistance, diameter, thickness and yield strength are not single-valued quantities, but also have distributions associated with them.

The acquisition, quality control and subsequent use of data is of paramount importance for the adequate execution of a Quantitative Risk Assessment. Therefore, British Gas has joined a Joint Industry Project on QRA, which has proved to be instrumental in improving data quality and quantity for subsequent work (DEA(E)-64).

2.2.4 QRA Using Randomly Distributed Variables

Once the input PDFs have been defined, it is then necessary to begin the design process. This, perhaps surprisingly, can be carried out in almost the same way as a standard design. The same equations may be used: the crucial difference is that the input variables have changed from assumed nominal values to more realistic random distributions.

QRA uses probabilistic mathematics and statistics to factor together the load and resistance variables into two distributions (Figure 2-4). The first defines all possible values that a load case can have. For a kick load case, this would cover the surface pressure experienced from 0.0955~0.1909m^3 condensate kick, right through to a very large gas kick at surface. The second distribution will govern casing resistance. The data space will cover the range in which all the input variables conspire to produce, for example, a very low collapse value (reduced yield strength, thinner wall thickness, ovality, etc.) to the equally unlikely situation where all those variables

combine to give the casing a very high collapse resistance. The result will be two opposing distributions: load and resistance.

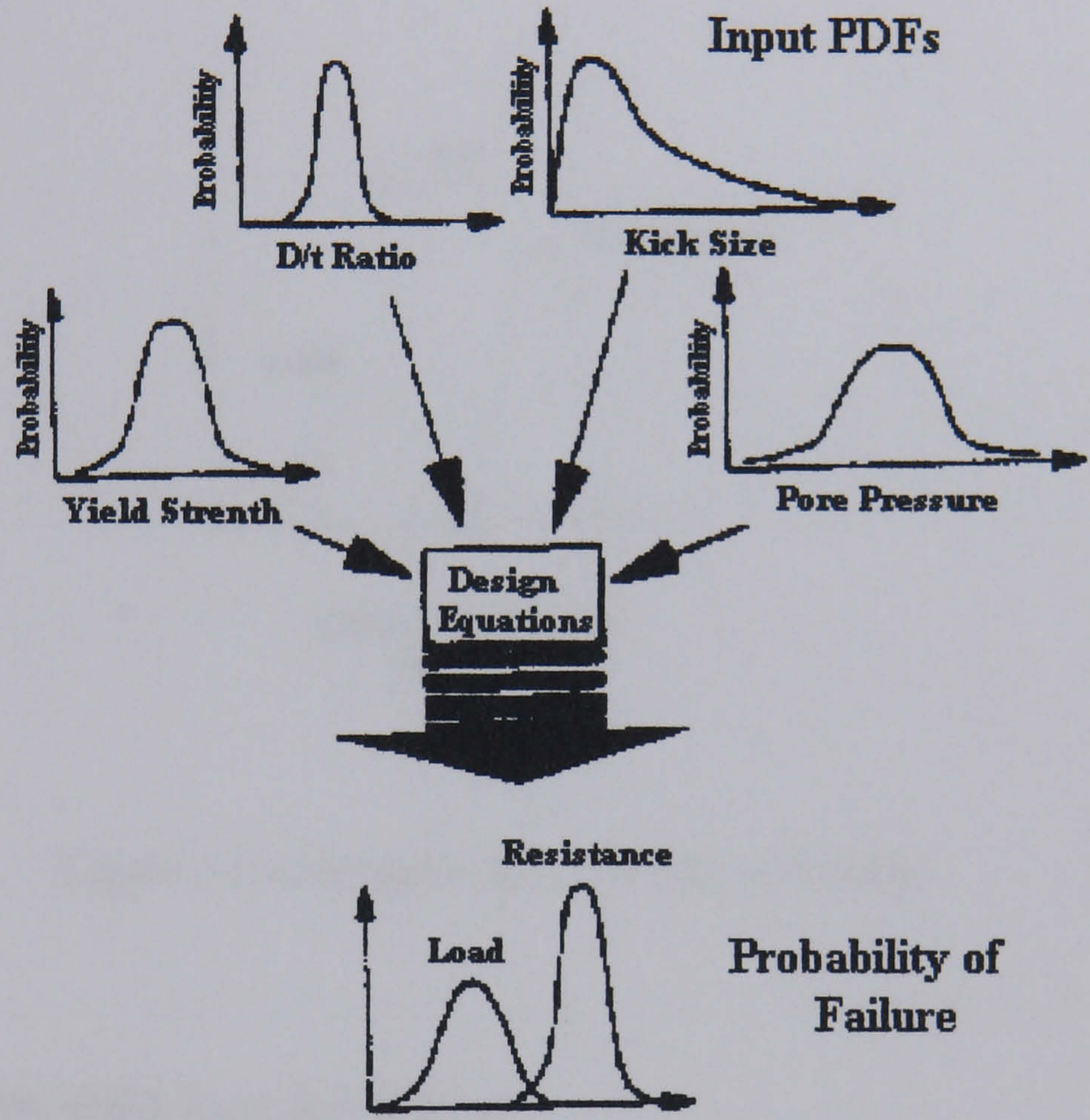


Figure 2-4: QRA design process

Figure 2-4 illustrates the point that despite a safety factor and good design practices, there will be a few times in a large number of applications of a particular design, in which load exceeds resistance. This is almost inevitable, but rather than ignoring the possibility, it is far more sensible to engineer the design to ensure that this failure rate is reduced to an economic level. This can be done by examining the casing

parameters governing resistance, and adjusting them to reduce the proportion of the data space for which the load exceeds the resistance. This relative area will determine failure rate (e.g. a value of 10^{-4} would mean one failure in every 10,000 applications).

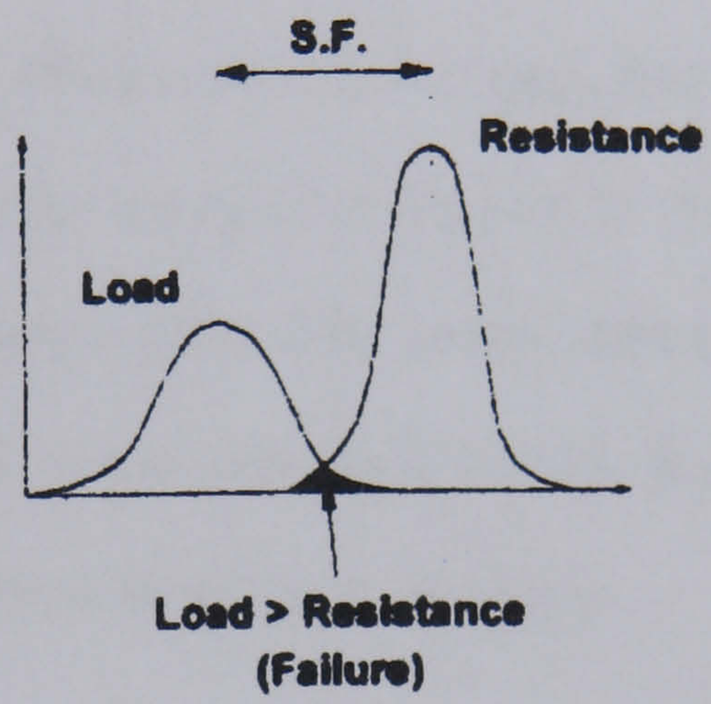


Figure 2-5: Determination of failure probability

2.2.5 Tolerable Risk Levels (TRLs)

Using QRA, it is possible to determine a failure rate for a casing design. This can then be compared to Tolerable Risk Level (i.e. maximum acceptable failure rate), and the casing parameters chosen to ensure that the failure rate is below this level.

Obviously, the risk of failure of any engineering design should be As Low As Reasonably Practicable (ALARP). The principle requires that every engineering design must have an ALARP failure rate, given current levels of technology and knowledge, as well as all reasonable expenditure to make the design safe.

Risk in an engineering sense can be defined as:

$$Risk = Failure\ Rate \times Consequences \tag{2-6}$$

The TRL thresholds identified in the report range from 10^{-4} to 10^{-6} (Marc Nunn, 1998). The higher bound is broadly acceptable for most operations, whereas a risk level towards the 10^{-4} end of the band must be investigated and, if possible, lowered.

Thus, QRA involves not only an assessment of the likely probability of failure, but also consideration of the effects (severity) of each type of failure. For example, a casing collapse at depth may not involve the cost or safety risks of a blow-out. It may be acceptable to assign a lower TRL to this failure type, say 10^{-4}, than for a blow-out. Non safety-risk events could be left to the discretion of the company, and based on the cost implications for this type of failure.

2.3 Summary

From the QRA approach discussed in the former sections, QRA is premised upon a set of design equations used. In this research, the design equations given by DEA (E)-64 are adopted.

The main aim of this research is to investigate the existing QRA methods and try to build up an accurate QRA simulation procedure for casing design. In the following chapters First/Second Order Reliability Method (FORM/SORM), Monte Carlo Simulation (MCS) will be discussed and a new methodology will be built. Unlike the traditional methods in which variables are often assumed to have a particular distribution (typically normal or log-normal), the new method uses Generalised Pareto Distribution (GPD) and important sampling (IS) techniques to obtain a good

fit on the tail area of the distribution. Because of the nature of the casing design there are normally very limited data available, it will be proved in chapter 6 that the new method can give more accurate results based on these limited random data points. A QRA toolkit will be developed at the end of this research for casing design using both the existing and the new methods.

Chapter 3 : QRA Methods

3.1 Introduction

There are several different methodologies available when adopting a QRA approach. The objective of this chapter is to review the traditional QRA methods such as First/Second Order Reliability Method (FORM/SORM) and Monte Carlo Simulation (MCS). These methods are the most frequently used in QRA.

3.2 FORM/SORM

First- and second-order reliability methods have been extensively applied to the probabilistic modelling of structural systems and, recently, to the oil and gas well drilling design. In this section, a description of the reliability formulation will be provided.

A scalar limit-state function, g(X), is formulated to define the probabilistic event, as a function of the random variables $X = (X_1, X_2, \cdots, X_n)$. These random variables can include the parameters such as outside diameter, wall thickness, ovality, etc. The limit-state function is formulated with the convention that the event of interest occurs when $g(x_1, x_2, \cdots, x_n) \leq 0$, where $x_1, x_2, \cdots, x_n$ are realisations of the random variables $X_1, X_2, \cdots, X_n$. In probabilistic risk assessment, the probability, that the estimated risk at a point of human exposure exceeds some regulatory standards, is more interested in. The corresponding g-function is given as:

$$g(X) = Risk_t - Risk_X \tag{3-1a}$$

where $Risk_t$ is the specified target risk, and $Risk_X$ is the estimated risk as a function of the random variables. In casing design, it can be written as follows,

$$g(X) = \text{Resistance} - \text{Load} \tag{3-1b}$$

The probability of exceeding a specific target risk level, termed the probability of failure is given by

$$P_F = P[g(X \le 0)] = P[Risk_X \ge Risk_t = \int_{g(X)\le 0} f_x(x)dx \tag{3-2}$$

where $f_x(x)$ is the joint Probability Density Function (PDF) of X and the integration is performed over the failure domain. The joint PDF of X is seldom known, and even in these cases the numerical integration of the above multi-fold integral is quite cumbersome. In light of these difficulties, the reliability methods provide an alternative means for approximating the above integral, along with the relevant sensitivity information.

In the reliability approach, the random variables and the limit-state function are transformed using a non-linear mapping to the standard normal space of uncorrelated normally distributed variable, u, of zero mean and unit variance. Then, the transformed limit-state surface, $G(u)=0$, is approximated at a point on the surface which is closest to the origin. This point is called the design point, and it is the most likely failure point in the standard normal space. This is followed by an estimation of a first- or second-order approximation of the probability of failure. The distance from the origin to the design point in the standard space is termed the reliability index, β, and is a measure of the reliability of the system considered.

The major advantage of the non-linear mapping is the rotational symmetry of the probability density in the standard normal space, which means that for all hyper-planes of equal distances from the origin, the probability content of the half space away from the origin is constant. Furthermore, the probability density in the standard normal space decays exponentially with the square of the distance from the origin. Therefore, the primary contribution to the probability integral in equation (3-2) comes from the part of the failure region closest to the origin. Consequently, the design point is an optimum point at which to approximate the limit-state surface G(u)=0.

The determination of the design point is obtained by posing the problem in a constrained non-linear optimisation formulation where the distance from the origin in the standard normal space is minimised, such that the point lies on the surface defined by G(u)=0.

In FORM, the approximation is carried out in the standard normal variable space $\boldsymbol{Y}$.

The transformation between $\boldsymbol{X}$ and $\boldsymbol{Y}$ is expressed by

$$\boldsymbol{Y} = \boldsymbol{T}(\boldsymbol{X}) \tag{3-3}$$

In the space of $\boldsymbol{Y}$, we write the corresponding limit-state function as

$$G(\boldsymbol{Y}) = G(\boldsymbol{T}(\boldsymbol{X})) = g(\boldsymbol{X}) \tag{3-4}$$

and then

$$P_f = \int_{G(y)\le 0} \phi(y)dy \tag{3-5}$$

where $\phi(y) = (2\pi)^{-n/2} \exp(-\frac{1}{2}\|y\|^2)$ is the standard normal density function with n denoting the number of random variables.

Based on special properties of the standard normal space, the FORM approximation is carried out by linearising the limit-state function $G(\boldsymbol{y})$ at the "design point" $\boldsymbol{y}^*$, which is the point on $G(\boldsymbol{y}) = 0$ nearest to the origin and is obtained as the solution to the constrained optimisation problem

$$\min\{\frac{1}{2}\|\boldsymbol{y}\|^2 | G(\boldsymbol{y}) = 0\} \tag{3-6}$$

Based on the linearised limit-state function, the first-order approximation to the probability of failure is

$$P_f \approx \Phi(-\beta) \tag{3-7}$$

where

$$\beta = -\frac{1}{\|\nabla G(\boldsymbol{y}^*)\|}\langle \boldsymbol{y}^*, \nabla G(\boldsymbol{y}^*)\rangle \tag{3-8}$$

here $\langle\cdot,\cdot\rangle$ denotes the inner product of vectors, is the first-order reliability index. The sensitivity of the reliability index to $\boldsymbol{y}^*$, denoted by $\boldsymbol{\alpha} = \{\frac{\partial\beta}{\partial y_1}, \frac{\partial\beta}{\partial y_2}, ..., \frac{\partial\beta}{\partial y_n}\}$, is given by

$$\boldsymbol{\alpha} = \text{sgn}(\beta)\frac{\boldsymbol{y}^*}{\|\boldsymbol{y}^*\|}, \tag{3-9}$$

where $\text{sgn}(\cdot)$ denotes the sign of the argument.

An accurate determination of the design point is essential for FORM analysis. Typically, a constrained optimisation algorithm is used that requires repeated evaluations of the limit-state function $g(\boldsymbol{x})$ and its gradient $\nabla g(\boldsymbol{x})$.

It is necessary to understand the limitations of the reliability methods since one cannot guarantee convergence to a global minimum. Nevertheless, different starting points may be used, and convergence to the same point is considered to indicate global convergence. Furthermore, the method could also become numerically intensive for problems characterised by very large numbers of random variables, where many function evaluations may be needed to numerically estimate the gradients required by the optimisation algorithm to determine the design point. Analytical estimation of the required gradients will alleviate this problem. The inaccuracy of FORM for highly non-linear problems can be solved, in many cases, by using SORM.

FORM and SORM differ mainly in their method of failure surface approximation in the standard space. In FORM, the limit-state surface is approximated by the tangent hyper-plane at the design point, the probability of failure only depends on the location of the design point. SORM, on the other hand, performs the approximation by fitting a paraboloid at the design point, it can provide more accurate results than FORM, especially when the limit state is parabolic about the design point.

It is therefore clear that the quality of the approximation will depend on the extent of nonlinearly of the limit state surface at the design point. In SORM, the principal curvatures of the limit-state surface at the design point are used to construct the paraboloid approximation of the surface, which is used to compute a second-order estimate of the failure probability. The principal curvatures are the eigenvalues of the Hessian (second-derivatives) matrix of the surface. This matrix can be obtained using the finite-difference method. SORM is therefore computationally more expensive

than FORM. The decision on which approximation method to use should be made by comparing the needed accuracy to available computational resources.

3.3 Monte Carlo Simulation

3.3.1 Introduction to Monte Carlo Simulation

As an alternative to using analytical procedures or First Order Reliability Methods (FORM) for solving probability based and reliability problems, a technique which has had a great impact in many different fields of computational science is a technique called Monte Carlo Simulation (MCS). This technique derives its name from the casinos in Monte Carlo - a Monte Carlo simulation uses random numbers to model some sort of a process. This technique works particularly well when the process is one where the underlying probabilities are known but the results are more difficult to determine. A great deal of the CPU time on some of the very fast computers is spent performing Monte Carlo simulations because we can write down some of the fundamental laws of physics but cannot analytically solve them for problems of interest.

Monte Carlo methods are computations on random numbers. This type of mathematics is experimental, to be contrasted with theoretical mathematics in much the same way as one distinguishes between, for example, experimental and theoretical physics. The essential difference between experiment and theory is that the experimenter infers results from observed data, whereas the theoretician deduces conclusions from assumed postulates. In Monte Carlo work the raw observational data are a set of random numbers; the way in which they are handled

computationally, to yield some desired result, constitutes the experimental technique. As in a physical experiment the observed data, and hence the results, are subject to experimental error. The better the experimental technique, the smaller are the subsequent errors in the final results. The real art of Monte Carlo work is to obtain an acceptably small error in these results without labouring over their production.

The expression Monte Carlo method is actually very general. Monte Carlo (MC) methods are stochastic techniques, which means they are based on the use of random numbers and probability statistics to investigate problems. You can find MC methods used in everything from economics to nuclear physics to regulating the flow of traffic. Of course the way they are applied varies widely from field to field, and there are dozens of subsets of MC even within chemistry. But, strictly speaking, to call something a Monte Carlo experiment, all you need to do is use random numbers to examine some problem.

3.3.2 Crude Monte Carlo simulation

As with FORM/SORM, it is first necessary to express the safety of reliability problem in terms of a safety margin equation as following:

$$M = g(X_1, X_2, \cdots, X_n) \tag{3-10}$$

where M is the random safety margin and X_i are the uncertain quantities(random variables) which may obey any form of probability distribution and which together govern whether or not failure occurs. As with FORM, $M \leq 0$ corresponds to failure. Each random variable has a corresponding probability density function $f_{X_i}(x_i)$ and distribution function $F_{X_i}(x_i)$.

In basic Monte Carlo analysis, the approach is to run a number of 'trials' in which q sets of n random variables $X_i(i=1,2,...,n)$ are randomly generated. For each of the q sets the value of the corresponding safety margin M_j is determined using Eq(3-10), e.g.

$$\begin{aligned} M_1 &= g(X_{11}, X_{12}, \cdots, X_{1n}) \\ M_2 &= g(X_{21}, X_{22}, \cdots, X_{2n}) \\ &\vdots \\ M_q &= g(X_{q1}, X_{q2}, \cdots, X_{qn}) \end{aligned} \tag{3-11}$$

If, in the q trials, there are q_f occasions on which $M_j \geq 0$, then the failure probability is given by

$$P_f \underset{q\to\infty}{=} q_f / q \tag{3-12}$$

in practice, for finite and reasonably large q, the ratio q_f/q cab be used as an estimate of P_f.

As an alternative to use Eq(3-10) which defines the safety margin in terms of the basic variables X_i, the safety margin may be expressed in terms of the corresponding unit standard normal variables Z_i by appropriate transformations, i.e.

$$F_{X_i}(X_i) = \Phi(Z_i) \tag{3-13a}$$

or

$$X_i = F_{X_i}^{-1}[\Phi(Z_j)] \tag{3-13b}$$

where $F_{X_i}^{-1}[\cdot]$ is the inverse distribution function of the random variable X_i and Φ is standard normal distribution function.

Then, we may write

$$M = g\{F_{X_1}^{-1}[\Phi(Z_1)], F_{X_2}^{-1}[\Phi(Z_2)],..., F_{X_n}^{-1}[\Phi(Z_n)]\} \tag{3-14a}$$

or, more simply

$$M = g'(Z_1, Z_2,..., Z_n) \tag{3-14b}$$

For any specific set of realisations of the Z variables (i.e. $z_1, z_2,..., z_n$) we may evaluate the specific value of the safety margin m as

$$m = g'(z_1, z_2,..., z_n) \tag{3-14c}$$

The Monte-Carlo simulation may now be carried out simply by generating a set of independent unit standard normal variables and substituting them in Eq(3-14c) and checking on whether or not $m \leq 0$.

The advantages of the simulation are:

(1) Once the safety margin has been established (remember this is an entirely deterministic concept), P_f can be evaluated very simply by repeated calculation of the values M_j.

(2) The accuracy of the estimate of P_f can be increased indefinitely by increasing the number of trials.

(3) The concept is very easy to understand.

Meanwhile, the disadvantages of the above approach are as follows:

(1) If P_f is very small, q_f may be zero unless the number of trials is very large. In this case P_f is undefined.

(2) Large numbers of trials may involve large computing effort, especially if the function $g(\cdot)$ requires complex calculations.

(3) The value of P_f obtained is itself a random variable (i.e. subject to statistical uncertainty). So by chance high or low estimate of P_f will be obtained in any set of m trials.

Detailed crude MCS Algorithm is given in the Appendix B. In next section, Monte Carlo Method for solving multivariable problems will be given.

Appendix H describes the algorithms of Monte Carlo Method for solving multivariable problems.

3.4 Summary

The traditional simulation methods of QRA, for example, FORM, SORM and MCS, have been discussed in this chapter. The advantages of these methods are that they give a simple procedure and its algorithms are easy to be implemented.

Because of the nature of FORM and SORM, FORM is accurate when the limit state is a linear hyper-plane and that the probability of failure only depends on the location of the design point. SORM can provide more accurate results than FORM, especially when the limit state is parabolic about the design point.

Monte Carlo Simulation is one of the most widely used methodologies for QRA because it is easy to program and apply.

However, the disadvantages of FORM and SORM are:

a) they depend upon precise knowledge of the distribution of the basic input variables which are often not known in practice; and

b) the design point can fall outside the range of experimentally measured data for the input variables.

The disadvantages of MCS are:

a) it is very computationally intensive when the simulated event is of very low probability;

b) is limited to the central part of the distribution of the basic input variables and observed data.

In casing design in the oil and gas industry, the failure probability is very small. That is to say, we are most interested in obtaining information that happens to fall in the tail of the distribution, rather than the centre of it. It can be shown that when simulating an event with a probability p, and in order to get a coefficient of variation of the ensemble of 10%, the necessary Monte Carlo sample size should be around $100/p$. Hence, if the probability of failure is very low, one may be required to execute tens or even hundreds of thousands of Monte Carlo simulations to obtain the results with acceptable accuracy.

The following chapters will try to build a model using Generalised Pareto Distribution (GPD) and important sampling (IS) techniques to obtain a good fit on the tail area of the distribution. Because of the nature of the casing design there are normally very limited data available, it will be proved in chapter 6 that the new method can give more accurate results based on these limited random data points. A QRA toolkit will be developed at the end of this research for casing design using both the existing and the new methods.

Chapter 4 : Distribution Tail Behaviours in Quantitative Risk Assessment

4.1 Introduction

In reliability analysis a probabilistic model has two components. One is the mathematical description of its properties, this is probability theory. The other one is the description of its relation with reality, i.e. the data and the observations, this is statistical inference. Most research up to date has focused on the problem of the mathematical description of the model, while the second aspect has been somewhat neglected. As we discussed in the former chapters, Monte Carlo Simulation method (MCS) is one of the most widely used methodologies for Quantitative Risk Assessment. It is easy to program and apply and amenable to analytical and numerical models. A major disadvantage, however, is that it is very computationally intensive when the simulated event is of very low probability. In QRA simulation procedure, most existing statistical estimation procedures, such as MCS and FORM/SORM, try to obtain a good fit of the theoretical distribution in the centre of the observed data, rather than in the tail.

In casing design, we are most interested in obtaining information that happens to fall in the tail of the distribution, rather than the centre. For instance, one is usually interested in the probability that the estimated risk exceeds some predetermined regulatory threshold level. Such a probability can be in the bulk of the distribution (probability of 0.1-0.9), but it can also exist at a very low value (one in a few

thousand). It can be shown that when simulating an event with a probability p, and in order to get a coefficient of variation of the ensemble of 10%, the necessary Monte Carlo sample size should be around 100/p. Hence, if the probability of failure is very low, one may be required to execute tens or even hundreds of thousands of Monte Carlo simulations to obtain the results with acceptable accuracy.

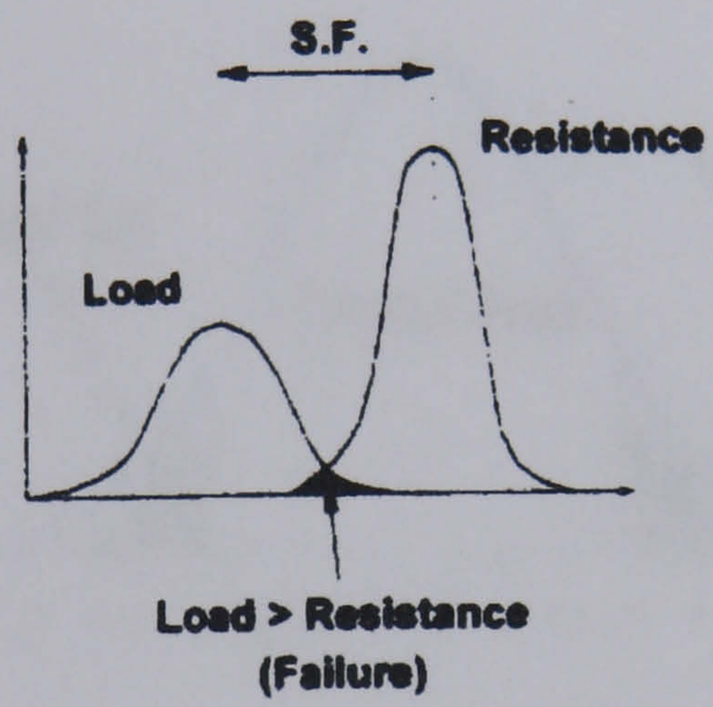


Figure 4-1: Failure area of QRA

The traditional simulation methods, such as Monte Carlo simulation, uses predefined the distribution type and runs the simulation over the whole distribution based on certain steps. As seen in Figure 4-1 the area which mostly effects the QRA result, lies in the "tail" area of the distribution. Whether we can get an accurate result or not critically depends on if we can get a good fit in the tail area of the variable distributions.

Figure 4-2 shows a sample of the distribution tail. The traditional methods do not have any distinction between the "central" area and the upper/lower "tail" area. It is

thus obvious that the "central" area does not make any effect to the results and it is a waste of time to run the simulation in this area.

In the following sections of this chaper will try to build a model which concentrates only on the tail area of the distribution and try to obtain a good fit on the tail to achieve maximum accuracy based on the limited data.

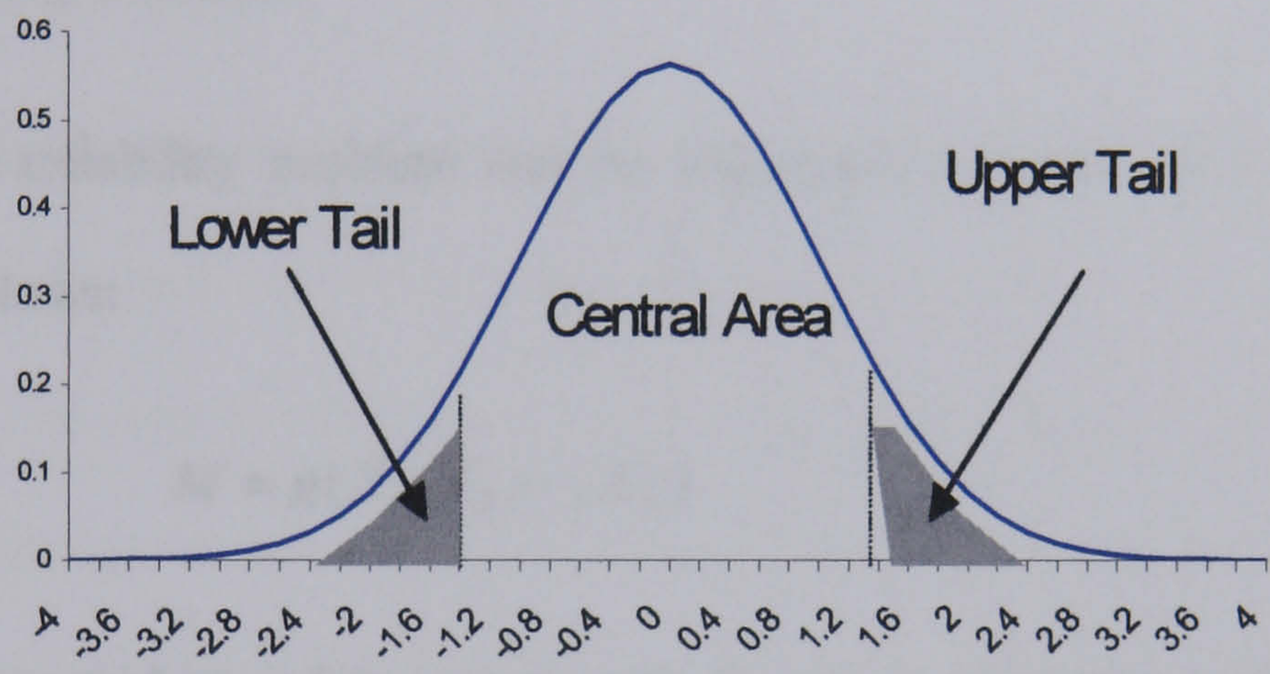


Figure 4-2: A Typical Distribution Tail

The aim of determining certain tail characteristics of random variables in the reliability-based design includes:

(1) the computation of high, or low, quantities of a random variable;

(2) the determination of very small probabilities of exceedance associated with fixed levels; or

(3) The accurate description of the tail of a probability distribution.

The latter objective is critically important in structural reliability analysis. If one fails to model the tail behaviour of the basic variables correctly, then the resulting reliability level is questionable.

In this chapter, a mathematical model will be built for tail estimation and will be contrasted with other simulation methods.

4.2 Probability Model

The safety of reliability problem can be expressed in terms of a safety margin equation as follows:

$$M = g(X_1, X_2, \cdots, X_n) \tag{4-1}$$

Where M is the random safety margin and X_i are n uncertain variables (random variables), which may obey any form of probability distribution and which together govern whether or not failure occurs.

With model-based risk or reliability analysis, risk is assessed by means of small probabilities associated with a random variable Z which arises as the response of a model M. The model can be deterministic or stochastic. It is a common characteristic of such model-based analyses that the tail area of the response variable Z is of interest, i.e. either the upper range or the lower range of this variable.

In what follows, it is assumed that risk is quantifiable through the probability of exceedance or non-exceedance of a set value by the variable Z. This small

probability or risk R_Z can consequently be computed on the basis of the known model M and known stochastic properties of the uncertainties X_i.

The principal interest now is to examine the change of R_z as a function of the assumed models for the tails of the basic uncertainty variables X_i.

4.3 Methods of Tail Estimation

Many statistical estimation procedures focus on estimating the central part of a probability distribution. In traditional statistical applications, the shape of the distribution in this region does indeed constitute the main subject of interest. This is, however, not the case in risk and reliability problems, where the interest often lies in the tail of a distribution. The use of a central method in conjunction with tail estimation amounts to estimate, for a given parametric distribution family, the values of the parameters (e.g. on the basis of maximum likelihood or a method of moments) and then using the tails of this estimated distribution in a subsequent analysis. As a result, the data in the central part of the distribution clearly have a dominating influence on the estimated distribution and the data in the tails which are usually few in number do not carry much weight. But, in order to get a good fit of the distribution tail, it is precisely these data that should have the most influence on the tail behaviour.

In reliability analysis the main modelling effort does not concern the central part of the distribution, but one or both tails of the distribution, since failure events are generally caused by extreme behaviour of random variables. The key concept on

modelling tail behaviour is that of tail equivalence. It gives a criterion for the quality of approximation of a distribution function F by another distribution G in the tail region. If we consider two distribution functions F(x) and G(x), these distribution functions are called tail equivalent if

$$\lim_{x\to\infty}\frac{1-F(x)}{1-G(x)}=1 \tag{4-2}$$

This means that, asymptotically, the relative error of approximating the small exceedance probability $1-F(x)$ by $1-G(x)$ approaches zero as $x\to\infty$.

As shown by Castillo (Castillo, E., 1988), the problem of approximating a distribution function in the tails is closely related to that of extreme values. The practical implication is that it becomes legitimate to fit an extreme value distribution to the tail of an arbitrary distribution, even though full asymptotic extreme value behaviour is not achieved. Actually, even though most probability distributions commonly used in engineering belong to the Gumbel domain of attraction for extremes, it is often the case that the sample size n is not large enough to detect the Gumbel trend. In these situations a penultimate (or, pre-asymptotic) approximation to extremes has been suggested (Pickands, J., 1975); examples abound in literature of Weibull and Frechet extreme value approximations, even though, strictly speaking, the Gumbel domain of attraction applies.

In order to capture this critical dependence of tail approximations on the sample size, n, while, at the same time, allowing for flexibility in the selection of the type of extreme value distribution, the best approach is to consider the generalised extreme value distribution (GEVD) defined by:

$$F(x \mid \theta) = \exp\left\{-\left[1-\frac{\gamma(x-\alpha)}{\beta}\right]^{\frac{1}{\gamma}}\right\} \tag{4-3}$$

The parameter set θ contains three values: β is a positive constant and

$$\begin{aligned} -\infty < x < \alpha + (\beta/\gamma) \quad & \text{When } \gamma > 0 \\ \alpha + (\beta/\gamma) < x < \infty \quad & \text{When } \gamma < 0 \end{aligned} \tag{4-4}$$

The problem is then reduced to finding the parameter values of α, β, γ, so that a good tail approximation is obtained. The criterion for good tail approximation will be discussed subsequently. If $\gamma = 0$, we have a type I or Gumbel distribution whereas the case $\gamma < 0$ corresponds to a type II or Frechet distribution and the case $\gamma > 0$ represents a type III or Weibull distribution.

4.4 A Methodology of Tail Model

4.4.1 The Aims and Basic Idea of Tail Model

In setting up a suitable tail fitting procedure, the following objectives should be accomplished:

- to avoid the problem of choosing cut-offs and discarding large parts of the data set;
- to assign greater weights to data in the tail, and less weight to data in the central and opposite parts of the data range;
- to respect the aforementioned principle of tail equivalence;

- to produce a method which is consistent with risk criteria used in structural reliability analysis and which can be interpreted visually;
- to produce estimates of the associated uncertainties.

By weighting the approximation error, which are larger in the distribution tail, it is ensured that the fit of the approximate distribution is good in the tail region. The selection of the weights is based on a risk criterion appropriate for structural reliability analysis.

4.4.2 Investigation of Generalised Pareto Distribution

To estimate the tail of a distribution in a more precise way, several methods have been proposed. A first method, by Maritz/Munro (Maritz, L.S. and Munro, A.H , March 1967)assumes that the whole distribution and not only the tail has an extreme value form. Functions of order statistics are then used to estimate the parameters of this distribution. Therefore, this method cannot be considered a proper tail fitting method.

Weissman (Weissman, I., 1978) and Boos (Boos, D.D., 1984) propose a method based on using only the higher order statistics to estimate the shape of the extreme value distribution. The problem here is that some relatively arbitrary cut-off number k must be determined from a sample of n data only the k largest are used. All the remaining sample information is discarded. A further extension of this method is given by Hasofer and Wang (Hasofer, A.M. and Wang, Z., 1991). They develop a test for the extreme value type based on these extreme order statistics. The method yields satisfactory results only for very large data sets, and quantile estimation is

possible only if the specified exceedance probability is greater then $1/n$, where n is the sample size.

The generalised Pareto Distribution (GPD) is a two-parameter distribution that contains the uniform, exponential, and Pareto distribution as special case. It has been used in wide variety of socio-economic applications and in failure problems of reliability studies (Hosking, J.R.M., et al. 1987, Smith, J.A., 1987, Van Montfort, et al. 1986). It is very useful to capture the asymptotic characteristics of the distribution tail.

The term generalised Pareto has also been associated in the statistical literature with a four-parameter distribution that is more general than the two-parameter form considered in present research. The two-parameter GPD is the one most used in engineering. It has been presented in some studies (Hosking, J.R.M. and Wallis, J.R., 1987) as the distribution of a random variable X obtained from the standard exponential distribution, Y, by a transformation of the form

$$X = \beta(1-e^{-kY})/k \tag{4-5}$$

X has the cumulative distribution function (cdf)

$$\begin{aligned} F(x) &= 1-(1-kx/\beta)^{1/k} && k \neq 0 \\ &= \beta^{-1}e^{-x/\beta} && k = 0 \end{aligned} \tag{4-6}$$

and probability density function (pdf)

$$\begin{aligned} f(x) &= \beta^{-1}(1-kx/\beta)^{\frac{1}{k}-1} && k \neq 0 \\ &= \beta^{-1}e^{-x/\beta} && k = 0 \end{aligned} \tag{4-7}$$

The range of x is $0 \le x < +\infty$ for $k \le 0$ and $0 \le x \le \beta / k$ for $k > 0$. β is a scale parameter and k is a shape parameter. One of the interesting features of this distribution is its simple mathematical form.

4.4.3 Tail heaviness index

It is important in the field of QRA to check whether an assumed distribution of an input variable X is indeed a fair and risk-consistent representation of reality. There is then a major difference between central models for which the use of classical statistical inference tools is appropriate, and tail models which are applicable to risk and reliability problems where the interest lies in the upper or lower tail area (Figure 4-2) (Pickands, J., 1975).

A useful parameter describing tail behaviour is the tail heaviness index (THI) (Breiman, L., Stone, C.J., and Gins, J.D., 1979, Hosking, J.R.M. and Wallis, J.R., 1987) which basically expresses how heavy any portion of a probability density tail is with respect to the pure exponential tail (THI=0). It was introduced by Breiman et al. (Breiman, L., Stone, C.J., and Gins, J.D., 1979) and first used by Boos (Boos, D.D., 1984) in a comparative study of techniques used to estimate large quantities. The idea is to benchmark heaviness against that of the exponential tail which is assigned a value of zero; it is negative for lighter than exponential tails (sub-exponential) and positive for heavier than exponential tails (super-exponential). The index is, generally, a function of the position on the tail, i.e. left-versus right-handed tail, as well as the exceedance probability q and the corresponding quantity

As described by Maes, M.A. and Breitung, K. (1993), for a value x on the lower/upper tail of the random variable X, it is defined as

$$\begin{aligned} THI(x) &= \frac{F(x)f'(x)}{f^2(x)} - 1 \qquad \text{Lower Tail} \\ THI(x) &= \frac{-(1-F(x))f'(x)}{f^2(x)} - 1 \qquad \text{Upper Tail} \end{aligned} \tag{4-8}$$

where F and f are the cumulative distribution function and the density of X, respectively. While for both types of tail the THI can most conveniently be related to the curvature of the minus-log-exceedance plot:

$$THI(x) = -\frac{L''(x)}{L'^2(x)} \tag{4-9}$$

where $L(x)$ is defined as:

$$\begin{aligned} L(x) &= -\ln(F(x)) \qquad \textit{Lower tail} \\ L(x) &= -\ln(1-F(x)) \qquad \textit{Upper tail} \end{aligned} \tag{4-10}$$

4.4.4 Tails Exceeding a High Threshold

It is clear from the Eq(4-9) that the tail heaviness index is proportional to the negative of the curvature $L(x)$ at any point x. Therefore there are three types of tails with negative, zero, and positive tail heaviness index. In a (L,x) plot, the exponential tail is represented by a straight line. A negative THI value points to a light tail of the beta-type, and most likely the existence of some upper bound and x. Conversely, a positive tail heaviness index corresponds to a concave (L,x) plot of a heavy tail of the Pareto-type.

Instead of using the concept of tail equivalence (4-2) to identify tail behaviour, we may also refer directly to the excesses $X-u$ of the random variable X over a high threshold u. Pickands (Pickands, J., 1975) showed that the Generalised Pareto Distribution (GPD) arises as the limiting distribution of the excesses provided the tail belongs to the domain of attraction of one of the extreme value distributions. The GPD can be written as :

$$
\begin{aligned}
F_{GPD}(x) &= \left[1+\xi\frac{(u-x)}{\sigma}\right]_+^{-1/\xi} \qquad (x<u) \qquad \text{Lower Tail} \\
F_{GPD}(x) &= 1-\left[1+\xi\frac{(u-x)}{\sigma}\right]_+^{-1/\xi} \qquad (x>u) \qquad \text{Upper Tail}
\end{aligned}
\tag{4-11}
$$

where u is the high threshold, and σ is a positive constant. As before, the case $\xi=0$ is interpreted as the limit $\xi\to 0$, which results in $F(x)=1-\exp(-\frac{(x-u)}{\sigma})$, i.e., the excess $X-u$ over the threshold u is an exponential random variable with mean σ. If $\xi>0$, Eq(4-8) represents one of several forms of the (unbounded) Pareto distribution, whereas the case $\xi<0$ restricts the range of the excess to the interval $0<X-u<-\frac{\sigma}{\xi}$, so that the GPD (4-11) becomes a (bounded) beta distribution.

The GPD enjoys widespread use in areas such as hydrology and oceanography. In fact, it forms a key component of peak-over-threshold models. It can easily be seen that a Poisson process of exceedances of a high level having excesses with a generalised Pareto distribution results in maximal which have a generalised extreme value distribution (4-3). Consequently, GEV and GPD are tail equivalent according to Eq(4-11) and the three types of extreme value distribution correspond directly to

the three tail-over-threshold. Moreover, it can be seen from the derivatives of the minus-log-exceedance function :

$$L_{GPD}(x \mid u,\xi,\sigma) = \frac{1}{\xi}\ln\left[1+\frac{\xi(x-u)}{\sigma}\right]_{+} \tag{4-12}$$

that the tail heaviness index of the GPD is constant over the entire range of x, and equal to:

$$THI_{GPD} = +\xi \quad \text{for all } x \tag{4-13}$$

The GPD represents the only family of distributions with constant heaviness.

Examples of light tails (i.e. negative heaviness index) include the normal, the gamma (with $\alpha > 1$), the logistic, the uniform (THI=-1), the beta, and most distributions with a finite upper range. Whereas some of the above eventually reach zero tail heaviness, ($\lim_{x\to\infty} THI = 0$) because they belong to the Gumbel domain of attraction of extreme values, their behaviour in the practical pre-asymptotic range is that of sub-exponential (beta-type) tails. Since the tail heaviness becomes more negative as x becomes smaller on the tail, this effect is more noticeable in applications with smaller sample sizes.

Positive THI values characterising long, heavy (Pareto-type) tails include the gamma (with $\alpha < 1$), the lognormal, the Pareto, the Cauchy, the Frechet. Similar to the gamma or the t-distribution, the Weibull distribution $F(x) = 1 - \exp\left(-\left(\frac{x}{u}\right)^{k}\right)$ has a tail heaviness index that can be either positive, zero, or negative before reaching zero at $x \to \infty$:

$$H_{weibull} = -(1-\frac{1}{k})\left(\frac{u}{x}\right)^{k} \tag{4-14}$$

Thus, for exponents $k > 1$, the Weibull tail is light, for $k = 1$ it is exponential, and for $K < 1$ the tail is heavy.

4.4.5 Char-square Tail Models

As introduced in the former sections, generalised Pareto Distribution (GPD) is a two-parameter distribution that contains the uniform, exponential, and Pareto distribution as special case, it has therefore been used in wide variety of socio-economic applications and in failure problems of reliability studies.

Several methods of fitting the GPD have been proposed, such as the method of maximum likelihood (ML), the classical method of moments (MM) and the method of probability weighted moments (PWM) (Greenwood, J.A. at al. 1979, Hosking, J.R.M., 1990). These methods have been compared in a simulation study by Hosking and Wallis (Hosking, J.R.M. et al. 1987), where it was shown that unless the sample size is 500 or more, estimators derived by the MM or PWM were more reliable than those obtained by the ML method. One disadvantage of the classical MM is that the order of the moments that it uses to estimate the parameters of a given distribution is somewhat arbitrary. It is well known, for example, that the classical use of the mean and variance, to fit a two-parameter distribution, is theoretically better defendable for certain types of distributions than for others.

Pickands (Pickands, J., 1975) showed that GPD is very useful when used to capture the asymptotic behaviour of the tail. Therefore, GPD is now used to model these tail

portions. For the ascending and descending parts of a distribution below the value u, the GPD distribution is given by Eq(4-8). Here ξ is exactly equal to the THI for both upper and lower tails.

Let the empirical distribution function of X be described by the n pairs (L_i, x_i) of the ordered sample $x_1 \le x_2 \le \cdots \le x_n$ in a (L,x) co-ordinate, where $L_i = -\ln\left(1-\frac{i}{n+1}\right)$. In the upper tail area, denote it by **T**, fitting a GPD distribution (4-8) which is tail-equivalent to the empirical distribution will be discussed.

The first problem is to define the extent of the upper tail area **T**, which amounts to determining the threshold u, the lower cut-off of **T**. The most useful formal tool is based on the equivalent of the mean residual life (MRL) diagram for lifetime modelling.

With the sequence of the ordered sample $x_1 \le x_2 \le \cdots \le x_n$ in a (L,x) co-ordinate.

Let $x_1 \le x_2 \le \cdots \le x_{(n_u)}$ be the observations that exceeds the threshold u, the MRL for the GPD model can be written as:

$$E(X-u \mid X>u) = \frac{\sigma+\xi u}{1-\xi}$$

where $\sigma + u\xi > 0$.

This is a plot of the conditional mean of the excesses $E(X-u \mid X>u)$ as a function of u. It provides an accessible approximation to the mean excess function.

The only stable part of this graph is located in the tail, and it often provides for a straightforward threshold selection. The slope and the intercept of the best MRL straight line fit can subsequently be related to the parameters ξ and σ of the GPD in the tail area. The selection of a lower tail threshold can also be avoided altogether by using a weighting technique as described by Maes, M.A. and Breitung, K. (1993).

Other methods for estimating the GPD parameters include maximum likelihood estimation (sufficiently unstable to be avoided), and Hosking and Wallis' method based on a simple extension of the MRL plot (Hosking, J.R.M. and Wallis, J.R., 1987). In reality, however, tail estimation is essentially a decision making problem, no single tail fit is satisfactory for all purposes and all possible different types of future usage. Rather, the criteria selected for achieving a suitable tail fit are critically dependent on what it is one wishes to achieve; they should be consistent with the measure of risk associated with the application affected by the tail behaviour.

The most flexible option for achieving this objective is weighted least squares on (L_i, x_i) in the tail region. The weights w_i depend on the quantity (or property) of interest so that, in general, the GPD parameters can be found from the minimisation with respect to ξ and σ of the sum of weighted square errors SWSE :

$$\text{SWSE} = \sum_{i \in T} w_i \left(L(x_i) - L_{GPD}(x_i \mid u, \xi, \sigma) \right)^2 \tag{4-15}$$

where L is given by Eq(4-12). This implies that errors $\Delta L = L_i - L$ are with mean 0 and variance proportional to $1/w_i$.

In order to minimise absolute errors on probability in tail area, supposing risk can be measured in terms of absolute errors Δq_i on small probabilities of exceedance q_i, we have,

$$\begin{aligned} -\Delta q_i &= \Delta F_i \\ &= (1-F_i)\Delta L_i \end{aligned}$$

it is clear that we need to the following weights in (4-15),

$$w_i = (1-F_i)^2 \tag{4-16}$$

Thus, an algorithm of the tail model based on optimising the GPD parameters can be achieved in the following steps:

- Estimate u on the basis of a mean residual life (MRL) diagram by identifying the stationary linear trend in the tail.

- Minimise the sum of weighted square error Eq(4-15) with respect to ξ and σ, where L is give by (4-15).

Obviously, the minimisation is consistent with the objective to obtain a risk estimate R_Z that is weighted correctly for tail effects.

It is of course very important to perform a complete uncertainty analysis for this estimation procedure. Uncertainty is associated first with the model given in (4-9) as well as with the parameters of the GPD themselves. Both can be treated in a way similar to that described by Castillo, E. (1988). This allows a basic tail modelling

uncertainty $\sigma_{L_i}^2(x)$ to be computed for the L_i value associated with each basic variable X_i.

4.5 Effect of Tails on QRA

The propagation of uncertainty through the model $g(X_1, X_2, \cdots, X_n)$ can be studied on a case by case basis. However, it can be shown (Gulati, D.L.Mckenna, et al. 1993) that in most practical cases, an error ΔL_i on the L function of the tail of the *i*th variable generates an error $\Delta(-\ln R_Z)$ on the logarithm of the risk R_Z approximately equal to:

$$\Delta(-\ln R_Z) \cong (1 + THI_i)\Delta L_i(x_{PLM}) \tag{4-17}$$

This approximation applies to asymptotic conditions, which are usually achieved when the risk R_Z is small and the number of variables is not excessive. The error on L_i is evaluated at the point of maximum likelihood (PML) associated with the model M. This is the point, which maximises the joint, log-likelihood of the variables $(X_1, X_2, \cdots X_n)$ subject to the constraint $g(X_1, X_2, \cdots X_n)$. This formula shows that the relative contribution of heavy tails (THI>0) is more pronounced that of light tails (THI<0), which includes, for instance, bounded tails.

The variance of the total uncertainty on $\ln R$, can now be determined as follows:

$$\sigma_{\ln R_Z}^2 = \sum_{i=1}^{n} (1 + THI_i)^2 \sigma_{L_i}^2 \tag{4-18}$$

The question of how the model-based risk R_Z varies as a function of the tail heaviness index of the *j*th uncertainty variable X_j can also be addressed.

4.6 Procedure of Tail model

We have discussed how Tail Model works based on the GPD model. In summary, the objective of the Tail Model is to find the GPD parameters, μ, ξ, and σ.

Here we first need to put the sample data set in order of $x_1 \le x_2 \le \cdots \le x_n$ and calculate in the (L, x) co-ordinate of n pairs (L_i, x_i), where

$$L_i = -\ln\left(1 - \frac{i}{n+1}\right)$$

Then, as described in section 4.4.5, MRL diagram can be used to estimate the threshold u by identifying the stationary linear trend in the tail.

With L given by Eq(4-12) the GPD parameters can then be obtained by minimising the SWSE described in Eq(4-15) with respect to ξ and σ.

4.7 Tail Model Validation

To investigate the characteristics of the Tail Model built in above, the following example was employed. It provides a comparison with traditional methods Monte Carlo Simulation (MCS), First/Second Order Reliability Method (FORM/SORM).

Example 1 is a simple case in which two independent variables are involved with experimental distributions. The reason of this simple case is because we can get the analytical results in this example and thus provide us with the ability to proof that the Tail Model can give close results.

Example 2 is also a theoretical example to investigate the effects of variable dimensions. It provides a comparison with traditional methods, Monte Carlo Simulation (MCS), First/Second Order Reliability Method (FORM/SORM), to prove that Tail Model can give more accurate results with small numbers of simulations.

Example 1:

In the simple case of two independent variables z_1 and z_2, with parameters μ and σ, the state function is as follows:

$$g(z_1, z_2) = \gamma^2 - z_1 z_2$$

For this simple example, we can get the exact results by reducing as follows,

$$P(F) = \Phi\left(-\sqrt{2}\frac{\ln\gamma - \mu}{\sigma}\right),$$

$$\frac{\partial P(F)}{\partial \mu} = \frac{1}{\sqrt{\pi}\sigma} e^{-\left(\frac{\ln\gamma-\mu}{\sigma}\right)^2}$$

$$\frac{\partial P(F)}{\partial \sigma} = \frac{1}{\sqrt{\pi}\sigma^2}(\ln\gamma - \mu) e^{-\left(\frac{\ln\gamma-\mu}{\sigma}\right)^2}$$

$$\frac{\partial P(F)}{\partial \gamma} = -\frac{1}{\sqrt{\pi}\sigma\gamma} e^{-\left(\frac{\ln\gamma-\mu}{\sigma}\right)^2}$$

These provide us the exact results given γ, μ, and σ, we can then use it to validate the new Tail Model.

The experimental distribution for z_1 and z_2 is log-normal distribution. The probability density functions has the following format:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2} \qquad 0 \le x \le \infty$$

Given mean and standard deviation, to use the GPD Tail Model, THI can be calculated from Eq(4-8) to Eq (4-10).

The simulation results are shown in Table 4-1 with a comparison to FORM, SORM and Monte Carlo Simulation.

Table 4-1 shows the comparison of the exact result and the one achieved by using Tail Model built in this chapter. It is clear that, using small quantity of iterations, a good result by using the new method can be achieved.

Table 4-1: comparison with exact results (γ=4.1548, μ=1, σ=0.150)

		$P(F)$	Error	$\partial P/\partial\mu$	Error	$\partial P/\partial\sigma$	Error	$\partial P/\partial\gamma$	Error
Exact result		3.167E-5	—	1.261E-3	—	3.569E-3	—	-3.304E-4	—
MCS(m=1,000,000)		3.179E-5	0.38%	1.276E-3	1.19%	3.614E-3	1.26%	-3.537E-4	7.1%
FORM		3.428E-5	8.20%	1.420E-3	12.6%	3.207E-3	10.1%	-2.953E-4	10.6%
SORM		3.250E-5	2.62%	1.316E-3	4.36%	3.372E-3	2.72%	-3.012E-4	8.84%
Tail Model	m =50	3.181E-5	0.44%	1.275E-3	1.11%	3.515E-3	1.50%	-3.010E-4	8.9%
	m =500	3.189E-5	0.69%	1.270E-3	0.71%	3.617E-3	1.34%	-3.018E-4	8.6%
	m =5000	3.170E-5	0.09%	1.264E-3	0.24%	3.552E-3	0.48%	-3.021E-4	8.5%

To further investigate the accuracy and efficiency of the Tail Model, we now choosing different values of γ in this example.

Figure 4-3 shows the results from FORM, SORM, MCS, Tail Model and also the expected results with γ increases from 1.0 to 5.0. Figure 4-4 gives the absolute errors of these results. It can be seen that, as γ increases, FORM becomes more accurate. This is because the limit state function has less curvature when γ increases. As γ decreases Tail Model becomes less accurate, this shows that it is only well suited to the tails. MCS provides very close results with 1,000,000 simulations.

Because FORM and SORM require only a small number of simulations they are both useful at the preliminary design stage. When more accurate results are required, the tail method is better than FORM and SORM, and more efficient than MCS.

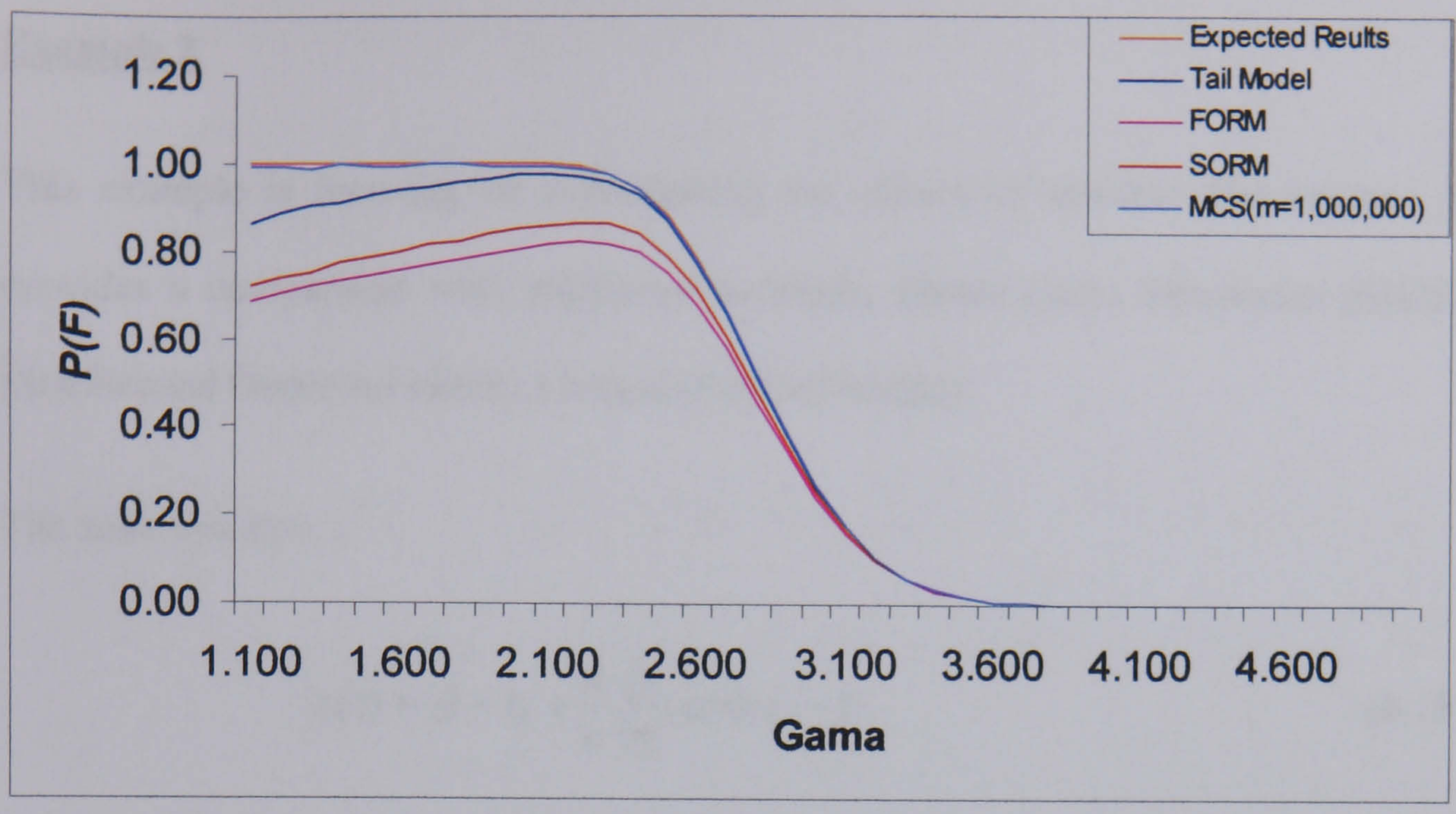


Figure 4-3: Probability results from different methods ($\mu = 1$ and $\sigma = 0.15$)

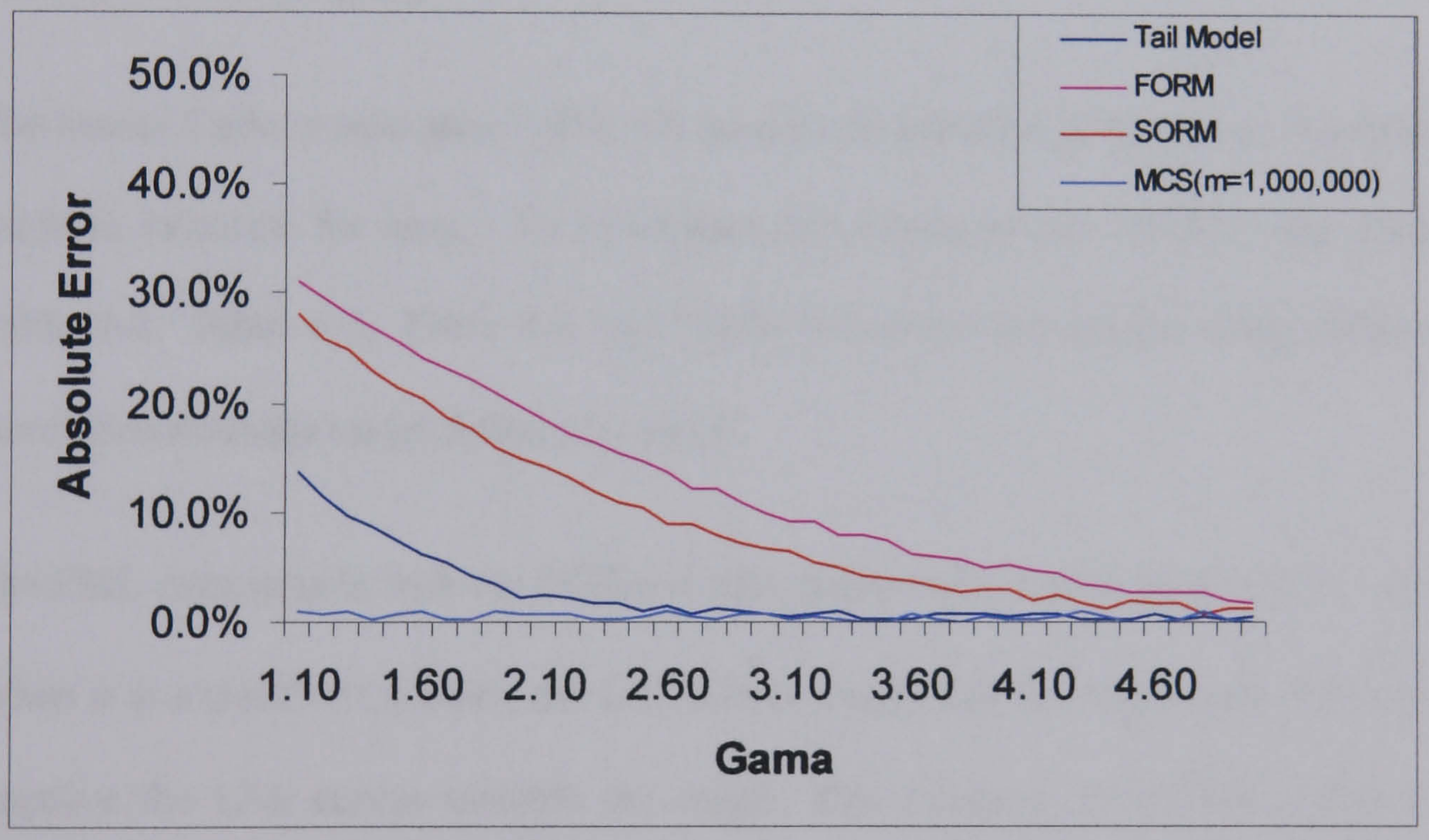


Figure 4-4: Error of the results with respect to the expected results

Example 2

This example is focusing on investigating the effects of variable dimensions. It provides a comparison with traditional methods, Monte Carlo Simulation (MCS), First/Second Order Reliability Method (FORM/SORM).

The state function:

$$g(z) = \beta - z_1 + \frac{a}{n}\sum_{i=2}^{n}(\cosh z_i - 1) \tag{4-18}$$

where $z_i (i = 1, \cdots, n)$ are n independent standard normally distributed random variables in interval [0,1].

The Monte Carlo results after 1,000,000 iteration is introduced here as an "accurate" result to calculate the error. To investigate the effects of the variable dimensions, Table 4-2, Table 4-3, Table 4-4 and Table 4-5 show the results using different simulation methods under different α and β.

The PML corresponds with the FORM design point and it is located at $(\beta, 0, \cdots, 0)^T$. When α is a positive constant, the LSS curves away from the origin and when it is negative the LSS curves towards the origin. The accuracy of SORM estimate is acceptable for low dimensionality. In contrast to FORM and SORM, Tail model provides a more robust means of determining a reasonably accurate probability of failure, even in the case of high dimensionality.

However, as the number of random variables increases, the accuracy and efficiency decreases rapidly for all methods considered. The errors using the Tail method are also excessively large.

Table 4-2: Comparison of different methods under different dimension (β=5, α=1)

		$n=2$	Error	$n=10$	Error	$n=50$	Error
MCS (m =1,000,000)		1.5120E-07	-	5.6500E-08	-	1.2500E-08	-
FORM		3.0100E-07	99.07%	3.0100E-07	432.74%	3.0100E-07	2308.00%
SORM		1.6045E-07	6.12%	7.0160E-08	24.18%	2.8370E-08	126.96%
Tail Model	m =50	1.6065E-07	6.25%	6.4037E-08	13.34%	1.7074E-08	36.59%
	m =500	1.5719E-07	3.96%	6.1432E-08	8.73%	1.5640E-08	25.12%
	m =5000	1.5344E-07	1.48%	5.9478E-08	5.27%	1.5008E-08	20.06%

Table 4-3: Comparison of different methods under different dimension (β=3.5, α=1)

		n =2	Error	n =10	Error	n =50	Error
MCS (m =1,000,000)		1.332E-4	-	4.229E-5	-	2.913E-5	-
FORM		2.326E-4	74%	2.326E-4	400%	2.326E-4	200%
SORM		1.403E-4	5.3%	6.028E-5	42%	4.434E-5	52%
Tail Model	m =50	1.410E-4	5.8%	4.750E-5	12.3%	1.965E-5	32.5%
	m =500	1.290E-4	3.1%	4.602E-5	8.8%	2.158E-5	25.9%
	m =5000	1.315E-4	1.3%	4.513E-5	6.7%	2.250E-5	22.7%

Table 4-4: Comparison of different methods under different dimension (β=3, α=-0.1)

		$n=2$	Error	$n=10$	Error	$n=50$	Error
MCS (m=1,000,000)		1.5200E-03	-	1.6496E-03	-	1.5200E-03	-
FORM		2.6700E-03	75.66%	2.6700E-03	61.86%	2.6700E-03	75.66%
SORM		1.6200E-03	6.58%	2.2800E-03	38.22%	3.2224E-03	112.00%
Tail Model	m=50	1.6086E-03	5.83%	1.8317E-03	11.04%	1.9863E-03	30.68%
	m=500	1.5705E-03	3.32%	1.7755E-03	7.63%	1.9178E-03	26.17%
	m=5000	1.5408E-03	1.37%	1.7159E-03	4.02%	1.8061E-03	18.82%

Table 4-5: Comparison of different methods under different dimension (β=3, α=0)

		$n=2$	Error	$n=10$	Error	$n=50$	Error
MCS (m=1,000,000)		1.3604E-01	-	1.3604E-01	-	1.3604E-01	-
FORM		1.3489E-01	0.84%	1.3489E-01	0.84%	1.3489E-01	0.84%
SORM		1.3513E-01	0.67%	1.3513E-01	0.67%	1.3513E-01	0.67%
Tail Model	m=50	1.3539E-01	0.48%	1.3539E-01	0.48%	1.3539E-01	0.48%
	m=500	1.3652E-01	0.35%	1.3652E-01	0.35%	1.3652E-01	0.35%
	m=5000	1.3629E-01	0.18%	1.3629E-01	0.18%	1.3629E-01	0.18%

4.8 Summary

Example 1 has shown the comparison of the exact result and the one achieved by using Tail Model. It can be seen that, as γ increases, FORM becomes more accurate. This is because the limit state function has less curvature when γ increases. Tail Model can give not only the correct, but also a more accurate answer than other methods. As γ decreases Tail Model becomes less accurate, this shows that it is only well suited to the tails. MCS provides very close results with 1,000,000 simulations. Because FORM and SORM require only a small number of simulations they are both useful at the preliminary design stage. When more accurate results are required, the tail method is better than FORM and SORM, and more efficient than MCS.

The effect of the random variable dimensions are investigated in Example 2. It shows that, for the range of examples considered, FORM provides a very poor result even if only two variables are involved. The accuracy of SORM estimate is acceptable for low dimensionality. Although in certain special cases FORM and SORM can be accurate and efficient, Tail model provides a more robust means of determining a reasonably accurate probability of failure, even in the case of high dimensionality.

One of the most important advantages of using Tail Model is that it requires very few iterations to obtain acceptable accuracy. However, as the number of random variables increases, the accuracy and efficiency decreases rapidly for all methods considered. The errors using the Tail method are also excessively large.

In casing design, a number of variables will appear in the design equation, such as outside diameter, wall thickness, ovality, eccentricity, etc. and could go up to more than 20 variables. Therefore, the Tail Model built in this chapter needs to be refined

so that it can be used into casing design. In the next chapter, a new technique, Asymptotic Important Sampling, will be introduced to refine the Tail Model.

Chapter 5 : Asymptotic Important Sampling Tail Model

5.1 Introduction

In the Quantitative Risk Assessment of casing design, the failure probability is normally very small. As discussed in the previous chapter, the Tail Model presented cannot give a satisfactory result when the number of variables becomes larger. Important Sampling (IS) techniques are investigated by A. Harbitz (1983), V. Bourgund et al. (1996), and R.E. Melchers (1989) and are found to be an efficient technique for multivariate integration especially when the probability is small. In QRA, almost all PDFs have these characteristics. Therefore, it would be efficient to use it in the Tail Model.

The key idea of IS technique is based on the careful selection of a sampling density for subsequent use in an importance sampling scheme (K. Breitung, 1997). The selection is based on theoretical consideration of the structure of the integration domain in the original variable space. Due to the asymptotic properties of the sampling densities, this technique becomes increasingly efficient (in terms of obtaining a given level of accuracy) as the failure probability becomes smaller.

The following sections will discuss how to build an Asymptotic Important Sampling model based on Tail Model. The quality of method will be illustrated using several examples in next chapter.

5.2 Importance sampling

As described by H.O.Madsen et al. (1986) and R.E. MelChers (1987), the basic random variables reliability problem consists of finding the probability that a set of n random variables Z, representing the physical and cognitive characteristics of a system, is contained within a failure domain $F = \{z : g(z) \leq 0\}$

$$P(F) = \int_F f_z(z)dz \tag{5-1}$$

where $f_z(.)$ is the joint probability density function (PDF) of Z, and $g : \Re^n \to \Re$ is the state function of the system. The surface G={z: g(z)=0} that separates the safe set F^c from $\{z : g(z) < 0\}$ is referred to as the limit state surface (LSS). In the following, a generalisation of this problem to a failure domain defined by r state functions $g_i : \Re^n \to \Re(i = 1, \cdots, r)$ will also be considered:

$$F = \bigcap_{i=1}^{r} \{z : g_i(z) \leq 0\} \tag{5-2}$$

In order to solve the integral (5-1) by statistical sampling (R.Y. Rubinstein, 1981), it is of utmost importance to control and to minimise the sampling error. The most frequently applied variance reduction scheme is that of importance sampling (R.E. MelChers, 1987), the use of which for structural reliability problems was originally suggested by M. Shinozuka (1983) and by A. Harbitz (1983). A new set of random variables X is introduced with PDF h_X. The functional relationship between z and x may, in general, be non-linear and not even one-to-one. However, in what follows, a transformation of points will be considered in failure domain with identical Jacobeans $J_{zx} = def[\partial z_i / \partial x_j]$ for each contributing disjoint set, such that

$$h_z(z(x)) = h_X(x)\left|J_{zx}\right|^{-1} \tag{5-3}$$

Consequently, the multivariate integral (5-1) can be written as:

$$P(F) = \int I(z \in F)\frac{f_z(z)}{h_z(z)}h_z(z)dz = \int I(z \in F)\frac{f_z(x)}{h_X(x)}\left|J_{zx}(x)\right|h_X(x)dx \tag{5-4}$$

with I denoting an indicator function equal to 1 for true arguments and zero otherwise. In other words, P(F) is determined as the expectation of a function w(x) with respect to $h_X(x)$:

$$P(F) = E_h[w(x)] \tag{5-5}$$

where

$$w(x) = \begin{cases} \left|J_{zx}(x)\right| f_z(x)/h_X(x) & if\ x \in F \\ 0 & elsewhere \end{cases} \tag{5-6}$$

The importance sampling PDF $h_X(x)$ must be nonzero over the failure domain, to ensure that the average $\overline{P}_m$ of randomly simulated values of the function $w(x_i)\quad (i = 1,\cdots,m)$ serves as an unbiased estimator of the failure probability:

$$\overline{P}_m = \frac{1}{m}\sum_{i=1}^{m} w(x_i) \tag{5-7}$$

with a mean and a variance equal to

$$E(\overline{P}_m) = P(F) \tag{5-8}$$

$$\mathrm{var}(\overline{P}_m) = \frac{1}{m}\mathrm{var}(w(x)) = \frac{1}{m}\int_{R^2}[w(x) - p(F)]^2 h_X(x)dx \tag{5-9}$$

As discussed in basic textbooks about Monte Carlo method (J.M. Hammersley et al.,1964, R.Y. Rubinstein, 1981), minimum variance is achieved if w is as constant as possible over the failure domain and zero elsewhere. In practice, there are two

conflicting objectives: the first one is to construct a density h_X in such a way that it closely mimics the behaviour of f_z within F, and the second one is to ensure that of h_X is a joint PDF from which random variable vectors can easily be generated, as will be shown subsequently, this conflict is not as serious as it seems, owing to the benefit drawn from the fact that $P(F)$ is small for type of problems considered.

5.3 Choice of the location of the importance sampling PDF

In the existing literature, the choice of a sampling PDF h_X has been approached in many different ways. In case of single failure function g, it's clear that the centre of mass h_X should be located near the point, or region, within F where the original density f_z is the largest. Although some of the earlier work related to MCS importance sampling was concerned strictly with problems formulated in standard normal space, there is, a priori, no good reason to move away from the original variable space. In fact, many desirable properties such as linearity are bound to be lost by transforming the constraints g_i into another variable space. The fact that MCS importance sampling can be used in the original variable space regardless of the type of random variables Z is a significant advantage of the proposed method (K. Breitung, 1997).

In the following, we will specifically assume that the log-likelihood function

$$l_z(z) = \ln f_z(z) \tag{5-10}$$

and the r failure $g_i(i=1,...,r)$ are such that the global maximum of l_Z with respect to the set F as given by (5-2) occurs in unique point z^*, and that, furthermore, this point of maximum likelihood (PML) lies somewhere on the boundary of F. It is worthwhile noting that if Z happens to be the standard normal space, then $l_z(z)$ is proportional to $-|z|^2$, such that the PML coincides with the point in F that lies closest to the origin, the so-called design point. In general there is no exact correspondence between the PML and the design point in the associated standard normal space; K. Breitung (1997), however, did show that, asymptotically, they must be equivalent.

Several situations may now be distinguished. The base case scenario arises when the point of maximum likelihood z^* is located on just one limit state surface, say G_1, and when there are no other points z on G_1, or on any of the other LSF $G_2, G_3, \cdots, G_r$ where a local maximum of l_Z is achieved. A more general case arises when the PML z^* coincides with the intersection of more than one, say $k(k \le r)$, limit state surface. Finally, there may be several additional, say s, points, $z^{(*)j}$, on the irregular boundary of F where local maxima occur, with $l(z^{(*)j}) < l(z^*)$, for j=1,...,s. each of these points in turn may be located on one LSS, of on the intersection of the disjoint neighbourhoods of the s+1 points $\{z^*, z^{(*)j}, j=1,\cdots,s\}$; the failure probabilities for each of these zones are estimated and added to the result in an estimate for P(F). Without lack of generality, we may therefore restrict ourselves to the problem of selecting a sampling density for the case of a unique PML that is situated either on just one LSS G_1, or else on the intersection point of k out of the r limit state surface $(2 \le k \le r)$.

5.4 Choice of the type of the sampling density

A comprehensive review of importance sampling methods is given by R.E. Melchers (1991). It states that many types of sampling densities h_X have been suggested by different authors. Some have been implemented in importance sampling software packages such as ISPUD, an acronym for Importance Sampling Procedure Using Design Points, etc. An obvious way to ensure that the sampling density is concentrated in the most likely area of the failure domain, is simply to shift the original PDF $f_z(z)$ in such a way that its mean coincides with the PML (A. Harbitz, 1983). An equally straightforward choice is that of a uniform density over a region enclosing the PML (M. Shinozuka, 1983). V. Bourgund and C.G. Bucher (1996) consider a normal PDF $h_X(x)$ having its mean at the PML. While no guidelines for the choice of the variance-covariance matrix are given, the latter authors use unit variances with satisfactory results. This technique is referred to as non-adaptive importance sampling. Precisely the static nature of the sampling density results in it being far from optimal for general use. It does not account for the characteristics of the joint pdf of the original variables, nor does it change with the shape of the limit state surface.

Karamchandani et al. (1985) introduced a method of dynamic importance sampling, which is also based on the idea that as sampling progresses, the knowledge of the failure domain increases. It yields good results even with a "bad" starting point, and the PML need not be computed. However, since the procedure starts with crude MCS, the initial sampling phase can be very extensive if the probability of failure is small.

Other interesting sampling schemes have been developed such as directional sampling, radial sampling and censored sampling (R.E. Melchers, 1989), as well as the use of kemel-methods (G.L. Ang, et al. 1989). The construction of "exact" sampling densities in standard normal space and their implementation in a conditional expectation simulation method come close to the asymptotic methods proposed in this chapter. Importance sampling is also used by M. Hohenbichler and R. Rackwitz (1998) with the objective of updating Breitung's second-order asymptotic result (K. Breitung, 1984).

5.5 Univariate asymptotic importance sampling densities

The problem of finding small probabilities associated with the random variable $g(Z)$ is intimately related to extreme value theory. To illustrate this point, consider the univariate case of a single random variable Z and the simple limit function $g(z) = z_0 - z$. In other words, $P(F) = P(Z > z_0)$ represents the exceedance probability of z_0. Assume that the behaviour of the continuous PDF f_z of Z in the vicinity of z_0 is regular, in the sense that it belongs to the domain of attraction of one of the three extreme value distributions (J.Galambos, 1978). This condition effectively restricts the application of the subsequent method to PDFs that do not contain ripples or other local irregularities as caused, for instance, by trigonometric noise factors, but the use of such distributions in engineering may be viewed as rather pathological. In what follows, the objective is to determine the ideal importance sampling density h_X. Since we want h_X to be nonzero only in the failure

domain, it is appropriate to introduce the co-ordinate x, with $z \to x = z - z_0$.

According to (5-6) and (5-9), the sampling error on $\overline{P}_m$ is minimised when:

$$h_X(x) = \begin{cases} \dfrac{f_z(z_0+x)}{P(F)} = \dfrac{f_z(z_0+x)}{1-F_Z(z_0)} & x > 0 \\ 0 & x \le 0 \end{cases} \tag{5-11}$$

or, in terms of the cumulative distribution function (CDF), $H_X(x)$:

$$\begin{aligned} H_X(x) &= 1 - P(Z > x + z_0 \mid Z > z_0) \\ &= 1 - \frac{1-F_Z(z_0+x)}{1-F_Z(z_0)}, \qquad x > 0 \end{aligned} \tag{5-12}$$

This means that h_X is equal to the original density truncated below z_0. The application of straightforward extreme value techniques provides good approximations for h_X as $P(F) \to 0$, i.e. as z_0 approaches either ∞ or the upper bound of Z. Three cases arise.

The first and most frequent situation arises when f_Z belongs to the Gumbel family. With only very few exceptions (J.Galambos, 1978), the right tail of f_Z is then unbounded. It is characterised by the fact that, as $z \to \infty$, the ratio of density to exceedance probability becomes indeterminate:

$$\text{as } z \to \infty: \quad \frac{f_Z}{1-F_Z} = -\frac{f_Z'}{f_Z} = -\frac{f_Z''}{f_Z'} = \cdots = c_1 > 0 \tag{5-13}$$

where c_1 is a positive constant, which may be infinity. It follows from (5-13) that the limit behaviour of the log-likelihood $l_Z(z) = \ln f_Z(z)$ and its derivatives $l_Z' = f'/f$, etc., is equal to:

$$\text{as } z \to \infty: \qquad l_Z \to -\infty, \qquad l_Z' \to -c_1, \qquad l_Z^{(i)} \to o(1), \quad i \ge 2$$

Using (5-13), the CDF H_X given by (5-12) can be approximated by:

$$H_X(x) \sim 1 - \frac{f_Z(z_0 + x)}{f_Z(z_0)} = 1 - \exp\left[l_Z(z_0 + x) - l_Z(z_0)\right] \tag{5-14}$$

A Taylor expansion about z_0 together with the identities (5-14), results in

$$\begin{aligned} H_X(x) &\sim 1 - \exp\left[l_Z'(z_0)x\right] \qquad x > 0 \\ &= 0 \qquad\qquad\qquad\qquad \text{elsewhere} \end{aligned} \tag{5-15}$$

Hence, the asymptotic importance sampling density for $P(F) \to 0$ (or $z_0 \to \infty$) is an exponential PDF with mean $-1/l_Z'(z_0)$.

In the rather exceptional event that the unbounded tail of f_Z is such that the ratio (5-13) becomes zero as $z \to \infty$, then f_Z may belong to the Frechet domain of attraction provided an exponent k can be found such that the exceedance probability decreases proportionally to z^{-k}.

The resulting conditions:

$$\text{as } z \to \infty: \qquad \frac{z^{-k}}{1 - F_Z} = -\frac{k(k+1)z^{-(k+2)}}{f_Z'} = \cdots = c_2 > 0 \tag{5-16}$$

show that in the Frechet case, the derivative of the log-likelihood tends to zero. The function $\left[-zl_Z'(z)\right]$, however, remains finite as $z \to \infty$:

$$-\lim_{z \to \infty}\left[zl_Z'(z)\right] = k + 1 \tag{5-17}$$

Equations (5-16) and (5-17) may now be applied to (5-12), yielding the following asymptotic expression for H_X, valid for x > 0:

$$H_X(x) \sim 1 - \left(1 + \frac{x}{z_0}\right)^{-k} \tag{5-18}$$

The asymptotic importance sampling PDF is therefore a Pareto PDF with exponent k and scale z_0; the exponent k may be retrieved from the derivative of the log-likelihood as (5-17). When (5-17) is substituted in (5-18), it becomes clear that if l'_Z does not tend to zero, which implies that $-zl'_Z(z) \to \infty$ (as in the case of the Gumbel type) then, by Euler's limit rule, the exponential PDF (5-15) re-emerges.

A more important and practical situation arises if Z has a finite upper bound, b. Various cut-off conditions may arise depending on the order of the first nonzero left-derivative of $F_Z(z)$ at $z = b$. In almost all practical cases the bounded PDF f_Z belongs to the domain of attraction of the Weibull extreme value distribution, which means that it is characterised by an exceedance probability which is of the order of $(b-z)^k$, where $k > 0$, so that:

$$\text{as } z \to b: \qquad \frac{(b-z)^{-k}}{1-F_Z} = \frac{k(b-z)^{k-1}}{f_Z} = \frac{k(k-1)(b-z)^{k-2}}{f'_Z} = \cdots = c_3 > 0 \qquad (5\text{-}19)$$

Consequently, the asymptotic expression for H_X given by (5-12) becomes:

$$H_X(x) \sim 1 - \frac{(b-x-z_0)^k}{(b-z_0)^k} = 1 - \left(1 - \frac{x}{b-z_0}\right)^k \qquad \text{For } 0 < x < (b-z_0) \qquad (5\text{-}20)$$

In view of (5-19) the Weibull exponent k may also be estimated from the derivative of the log-likelihood function:

$$\text{as } z \to b: \quad k \sim -(b-z)l'_Z(z) + 1 \qquad (5\text{-}21)$$

An alternative expression for the CDF (5-14) is therefore:

$$H_X(x) \sim 1 - \left(1 - \frac{x}{b-z_0}\right)^{1-(b-z_0)l'_Z(z_0)}, \qquad 0 < x < b - z_0 \qquad (5\text{-}22)$$

Clearly, the ideal importance sampling density is now a Beta PDF over the interval $[0, b-z_0]$ with parameters $\alpha_1 = 1$ and $\alpha_2 = k = 1-(b-z_0)l'_Z(z_0)$. Again, it should be noted that if the upper bound b is large with respect to the PML z_0 then the above beta PDF transforms into the familiar exponential PDF (5-15):

$$\lim_{(b-z_0)\to\infty} H_X(x) = 1 - \lim\left(1 - \frac{x}{b-z_0}\right)^{1-(b-z_0)l'_Z(z_0)} = 1 - \exp(l'_Z(z_0)x), \qquad x > 0 \quad (5\text{-}23)$$

In order to shed more light on the quality of these asymptotic densities, it is very instructive to consider a range of applications. For instance, we may calculate $w(x)$ using (5-5), as well as the variance (5-8), and investigate the sensitivity of these quantities to change in the location and the parameters of the sampling density.

Two points should be noted from the discussion of the univariate case: the importance of the gradient of the log-likelihood ∇l at the PML, and the significance of the exponential PDF. Both of these aspects also play a prominent role in the multivariate case.

5.6 The basic concept of asymptotic approximations for failure probability

The basic concept of asymptotic approximations is described using Laplace-type integrals of the form

$$I(\beta) = \int_F e^{\beta^2 l(z)} dz$$

where F is the failure domain: $F = \{z : g(z) \le 0\}$. In the case of failure probabilities, where $l(z)$ is scaled log-likelihood function of a probability density,

there are some characteristics. Since in general the probability content of the failure domain is small and far away for the region where the PDF is large, the maximum of $l(z)$ is at some points of the boundary G of F. For such a maximum of function on a surface G, the Lagrange multiplier theorem gives some results about the structure of the Taylor expansion of l and g at the Point of maximum likelihood (PML) z^*. The gradients $\nabla l(z^*)$ and $\nabla g(z^*)$ are orthogonal to the tangential space of G at the PML and therefore

$$\nabla f(z^*) = \lambda \nabla g(z^*)$$

Further, the first directional derivatives of f and g in the direction of the tangential space vanish. So, for a Taylor expansion of l at z^* in the region of F near this point, the directional derivative expansion of l in the domain F, which is given by $-|\nabla l|$, is needed firstly. Further the second directional derivatives into the direction of the tangential space of G is needed at the PML of l and g to determine the behaviour of the function l on the surface G near z^*. With this information it's possible to derive the asymptotic approximation.

5.7 The basic multivariate asymptotic importance sampling density

The case of one limit state function g and a unique PML, z^* is considered first. Together with $g(z^*) = 0$ and $\nabla g \neq 0$, it is also assumed that the maximum is regular with respect to F, i.e. $\det C \neq 0$, where the $n-1$ by $n-1$ matrix C is equal to

$$C = -\left\{ l_{ij} - \frac{|\nabla l|}{|\nabla g|} g_{ij} \right\}_{i,j=2,\cdots,n} \tag{5-24}$$

and where double subscripts denote second order derivatives. This matrix can easily be recognised as the negative of the first cofactor of the Hessian of the Lagrangian function associated with the constrained optimisation problem - maximise the log-likelihood $l(z)$ subject to $g(z)=0$. The Hessian is available directly only for the particular case that all the constraints variety of techniques may be written as explicit functions of z. In a more general context, a variety of techniques may be used to solve the optimisation problem and simultaneously reconstruct the Hessian, or its inverse. The principal curvatures of the LSS are determined iteratively on the basis of information collected during the final steps of a gradient projection search.

By a simple and suitable translation and rotation we can introduce a new curvilinear co-ordinate system $x_1, x_2, \cdots, x_n$, such that:

1. the PML lies at origin, that is, $x^* = (0, \cdots, 0)$;

2. The unit vector $\hat{x}_1$ coincides with the direction of steepest descent at the PML:

$$\hat{x}_1 = -\left[\frac{g}{|\nabla g|}\right]_{z=z'} \tag{5-25}$$

3. The remaining $n-1$ co-ordinates $x_2, \cdots, x_n$ are surface co-ordinates in the direction of the main curvatures on the limit state surface at PML.

In practice, a third system of co-ordinates $y_1, y_2, \cdots, y_n$ needs to be introduced. This co-ordinate system is Cartesian with y_1 coinciding with x_1 and $y_2, \cdots, y_n$ being the projections of the $x_2, \cdots, x_n$ surface co-ordinate axes on the tangent hyper-plane at the PML.

By the above procedures, the basic problem is essentially similar to the univariate problem. An asymptotic approximation is to be found of the joint density $h_X(x)$ of the random vector X. Using key results from differential geometry and Laplace integrals, K. Breitung (1997) developed the foundations for the following important result, for small P(F), approximately:

$$h_X(x) \sim \exp\left\{-|\nabla l| x_1 + \frac{1}{2}\sum_{i,j=2}^{n}\left(l_{i,j} - \frac{|\nabla l|}{|\nabla g|} g_{ij}\right) xi, x_j\right\}, \qquad x_1 > 0 \qquad \textbf{(5-26)}$$

This result implies first of all that the random variable X_1 is stochastically independent from the other $(n-1)$ variables and that it has approximately an exponential PDF with mean $1/|\nabla l|$.

Note that since x_1 coincides with the direction of steepest descent, $|\nabla l| = -(\partial l / \partial x_1)_x$. which is equivalent to the one-dimensional expression obtained under (5-15). The remaining variables matrix $X_2, \cdots, X_n$ are seen to have a multivariate normal PDF with zero mean vector and covariance matrix $+C^{-1}$, where C is given by (5-24) with all the derivatives taken with respect to y. The parameters of the asymptotic importance sampling PDF are thus seen to be governed by

- ❖ the gradient of the log-likelihood at the PML,
- ❖ the curvatures of the log-likelihood function at the PML, and
- ❖ the curvatures of boundary of the failure domain at PML.

Note that the occurrence of a regular maximum at the PML, guaranteeing the non-singularity of C, implies that none of the variances of $X_2, \cdots, X_n$ become infinite. Because of the choice of co-ordinates there are no mixed second derivatives of g, i.e.

$g_{ij} = 0 (i \neq j)$. If , in addition, there would be no mixed second derivatives of h, then the Hessian of the Lagrangian C is diagonal. After dividing by $|\nabla l|$, one obtains the differences of the curvatures k_{li} and k_{gi} of the level surface log-likelihood and of the limit state surface; a particular diagonal element on row i of C would therefore be equal to $|\nabla l|(+k_{li} - k_{gi})$. Note that k_{li} cannot possibly be less than k_{gi}, else there would be no maximum at the PML. The case of a limit state hyper-plane (all $g_{ij} = 0$) is also specifically considered by K. Breitung (1997). In this case the $\{x\}$ co-ordinate system is itself Cartesian. Finally, it should be kept in mind that (5-26) is only an approximation which is good in an asymptotic sense. Its quality depends on the overall structure of the LSS and the log-likelihood function. But for small $P(F)$, the use of (5-26) is guaranteed to result in a small sampling error.

5.8 Implementation of AIS Tail Model

In the original variable space, a simple linear transformation links z to the local Cartesian co-ordinates y at the PML. The transformation between y and x, however, requires additional attention. The practical implementation of (5-26) rest on the following processes.

At the PML(point of the maximum likelihood), y=0, the importance sampling density (5-28) is determined, the normalising constant being $|\nabla l|(2\pi)^{-(n-1)/2}(\det C)^{1/2}$. A random sample $\{x_j\}$ (j=1,…,m) is generated, the first component being exponential with mean $|\nabla l|^{-1}$ and the (n-1) remaining components normal with mean

zero and covariance matrix C^{-1}. Let us concentrate on a random point S thus obtained in the failure domain. First suppose that the limit state surface coincides exactly with its second-order approximation, such that the co-ordinated $x_2, x_3, \cdots, x_n$ represent arc lengths on the paraboloid; they are the surface co-ordinates of the projection of A along the direction (5-27) onto the limit state surface. It follows that, in this case:

$$\begin{aligned} y_1 &= -\frac{1}{2}\sum_{i=2}^{n} k_{gi} y_i^2 \\ y_i &= \frac{1}{k_{gi}}\eta^{-1}(k_{gi}x_i), \quad i = 2, \cdots, n \end{aligned} \tag{5-27}$$

where k_{gi} are the principal curvatures in the directions $i(i = 1,2,\cdots,n)$. The function $a = \eta^{-1}(b)$ represents a numerically suitable inverse function of the non-dimensional arc length function $b = \eta(a)$ for a parabola:

$$\eta(a) = \frac{a}{2}\sqrt{1+a^2} + \frac{1}{2}\ln(a + \sqrt{1+a^2}), \qquad a \in \Re \tag{5-28}$$

However, the g-value of the point $(x_1 = 0, x_2, x_3, \cdots, x_n)$ may not be exactly equal to zero, in which case a small correction ε is needed to project this point back onto the true LSS, in the direction given by (5-25). This can be achieved by a single Newton correction using either the true gradient or its first order approximation with respect to $\nabla g(0)$. Alternatively, the type of application may justify the use of a full line search in the direction of x_1. In practice, the correction is quite small, except, for instance, in such case when there is close osculation between the level g surface and the level log-likelihood surfaces in the vicinity of the PML. ($k_{li} - k_{gl} \approx 0$ leads to a nearly singular C matrix in (5-26) and a large variance in the direction i.)

In the general case then, the following non-linear transformation $x \to y$ is applicable:

$$\begin{aligned} y_1 &= +x_1 - \frac{1}{2}\sum_{i=2}^{n} k_{gi} y_i^2 + \varepsilon \\ y_i &= \frac{1}{k_{gi}} \eta^{-1}(k_{gi} x_i), \quad i = 2, \cdots, n \end{aligned} \tag{5-29}$$

It is clear that since the correction ε does not affect any of the surface co-ordinates, the Jacobian of the transformation is independent of ε. Substituting the derivatives of the arc length function (5-28) and considering the additional fact that the Jacobian of the orthogonal transformation $z \to y$ is equal to one, it can be verified that:

$$\left|J_{zx}(x)\right| = \left|J_{zy}(x) J_{yx}(x)\right| = \prod_{i=2}^{n} [1 + (k_{gi} y_i(x_i))^2]^{-1/2} \tag{5-30}$$

where $y_i(x_i)$ is given by the Eq(5-29). It is now a sample matter to determine the value of the sample statistics $[w(x)]_j$ associated with each vector x_j (j=1,…,m) using Equation (5-6), (5-26) and (5-30).

5.9 Additional asymptotic sampling densities

The previous results are now extended to the more general failure domain F given by (5-2). As indicated before, the present research focuses on the case where the PML is located at the intersection of k limit states ($k \le 0$), such that:

$$\begin{cases} g_1(z^*) = \cdots = g_k(z^*) = 0 \\ g_j(z^*) < 0, \quad for \ j = k+1, ..., r \end{cases} \tag{5-31}$$

the new co-ordinates system with origin at the PML is defined as follows. The first k co-ordinates $x_1,...,x_k$ are the projections of $z-z^*$ on to the negatives of the normalised gradient of the active constraints:

$$x_s = -(z-z^*)\cdot\nabla g_s / |\nabla g_s|, \qquad s=1,...,k \tag{5-32}$$

It must be noted that, due to the Lagrange multiplier theorem, the gradient of l can be written in the form:

$$\nabla l = \sum_{s=1}^{k} \gamma_s \nabla g_s \tag{5-33}$$

we assume that the maximum is regular, i.e. $\gamma_s > 0$ for $s=1,\cdots,k$. The remaining $(n-k)$ local coordinates, $x_{k+1},\cdots,x_n$, are the arc lengths in the directions of the main curvatures of the surface $g_1(z)=g_2(z)=\cdots=g_k(z)=0$ at the PML.

The asymptotic form of PDF near the PML is now the product of a (r-k)-dimensional normal density with covariance matrix C^{-1}, where

$$C=\{c_{ij}\}_{i,j=k+1,\cdots,n} = -\left\{ l_{ij} - \sum_{s=1}^{k} \gamma_s g_{ij,s} \right\}_{i,j=k+1,\cdots,n} \tag{5-34}$$

(with the subscripts ij denoting second order derivatives at the PML with respect to the local Cartesian co-ordinates), and a k-dimensional exponential distribution with independent marginal having a mean $(|\nabla g_s|\gamma_s)^{-1}$. The derivation is similar to the case of one PML. The asymptotically optimal importance sampling density is then equal to:

$$h_Y(x) = \det C \left[\prod_{s=1}^{k} |\nabla g_s|\gamma_s \right] (2\pi)^{\frac{n-k}{2}} \exp\left\{ -\sum_{s=1}^{k} |\nabla g_s|\gamma_s x_s - \frac{1}{2} \sum_{i,j=k+1}^{n} c_{ij} x_i x_j \right\} \tag{5-35}$$

The non-orthogonality of the first k unit vectors must carefully be taken into account when determining the Jacobean J_{zx} in (5-6).

Finally, it is worthwhile to revisit the case of a density that is bounded in the neighbourhood of the PML, as was done above in the uni-variate case. If it is assumed the PML to lie on just one LSS ($k=1$), and if, after transformation the variable X_1 has an upper bound c, then the asymptotic importance sampling PDF can be expressed as follows:

$$h_X(x) \propto (1-x_1/c)^{|\nabla l|c} \exp\left\{-\sum_{s=1}^{k}\frac{1}{2}\sum_{i,j=k+1}^{n}c_{ij}x_i x_j\right\} \quad 0<x_1<c \tag{5-36}$$

This result can easily be obtained by noting that the transformation $u_1=-\ln(1-x_1/c)$ yields a set of variables $(u_1,u_2,\cdots,x_n)$ that have the joint PDF (5-26); the expressions for ∇l, ∇g and the density itself can subsequently be transformed to produce (5-36).

5.10 Parameter sensitivity

In the context of a structure reliability analysis, it is very useful to determine sensitivity or importance measures for failure probabilities (P. Bjerager, 1990). They can be obtained easily and with little additional effort using asymptotic importance sampling, the only requirement is to keep track of a score function involving the known gradient of the original log-likelihood function of the gradient of the state function.

First consider the case where θ is a distributive parameter. For simplicity, only one parameter is considered, the extension to a vector of parameter is straightforward.

The sensitivity of $P_\theta(F)$ to a small change of the parameter θ on which the joint density $f_x(z,\theta)$ is dependent, can be represented by the partial derivative:

$$\frac{\partial P_\theta(F)}{\partial\theta} = \frac{\partial}{\partial\theta}\left[\int_F f_z(z,\theta)dz\right] \tag{5-37}$$

which may be transformed as follows:

$$\frac{\partial P_\theta(F)}{\partial\theta} = \int_F \frac{\partial f_z}{\partial\theta}dz = \int_F \frac{\partial \ln f_z}{\partial\theta} f_z dz = \int w(x,\theta)\frac{\partial l(x,\theta)}{\partial\theta} h_X(x)dx = E_h\left[w\frac{\partial l}{\partial\theta}\right] \tag{5-38}$$

where $w(x,\theta)$ id defined as in (5-6). It is clear that the sensitivity can be estimated from the same random sample $\{x\}_j$ $(j=1,...,m)$ used for the calculation of P(F), by assembling the statistic $1/m\sum^m w(x_j,\theta)\partial l(x_j,\theta)/\partial\theta$.

The sensitivity of P(F) to state function parameter can be determined in a similar situation. We restrict the discussion to the case of one parameter τ and one state function $g(z,\tau)$. The sensitivity can then be expressed as:

$$\frac{\partial P_\tau(F)}{\partial\tau} = -\int_{G(\tau)} \frac{\partial g(z,\tau)}{\partial\tau}\frac{f_z(z)}{|\nabla g|}ds_G(z) \tag{5-39}$$

where $G(\tau)$ is the LSS $\{z: g(z,\tau)=0\}$ and where $ds_G(z)$ denotes surface integration over $G(\tau)$. The choice of local co-ordinates in AIS is precisely such that the n-1 co-ordinates $x_2, x_3, \cdots, x_n$ the principle directions of the LSS at the PML.

Consequently, AIS lends itself well to the calculation of (5-39) at very little expense within the same simulation needed to estimate the failure probability. This becomes clear when (5-39) is viewed as an expectation with respect to the $(n-1)$ dimensional joint density $h^*_{X^*}(x)$ of the variables $X^* = \{X_2, X_3, \cdots, X_n\}$ given by (5-26) with $x_1 = 0$ (in local co-ordinates the equation of the surface G is $x_1 = 0$):

$$\frac{\partial P_\tau(F)}{\partial \tau} = -\int_{R^{n-1}} \frac{\partial g(z,\tau)}{\partial \tau} \frac{f_z(z)}{|\nabla g|} |J_{zx}| dx^* \\ = -\int \frac{f_z}{h_X^*} \frac{\partial g}{\partial \tau} \frac{|J_{zx}|}{|\nabla g|} dx^* = E_{h^*}\left[-\frac{f_z}{h_X^*} \frac{\partial g}{\partial \tau} \frac{|J_{zx}|}{|\nabla g|} \right] \tag{5-40}$$

This expected value can easily be estimated by averaging the statistic in square brackets generated by the original random samples $\{x\}_j$ with the first co-ordinate x_1 put equal to zero. Expressions similar to (5-40) can be derived easily to determine parameter sensitivity in the case of multiple state functions.

5.11 Procedure of AIS Tail model

The random variables reliability problem consists of finding the probability that a set of n random variables Z, representing all variables involved in the design equation, is contained within a failure domain, i.e. $F = \{z : g(z) \le 0\}$. In order to solve the integral (5-1) by statistical sampling, it is of utmost importance to control and to minimise the sampling error.

The first step of AIS Tail Model is to find the new set of random variables X which represents the most critical points in the failure domain. The PDF of the new set of random variables is expressed as h_X in Eq(5-3) and Eq(5-26), it must be nonzero over the failure domain.

The implementation is discussed in section 5.8. $y_i(x_i)$ can be calculated from Eq(5-29), then the Jacobians $J_{zx} = \det[\partial z_i / \partial x_j]$ can be determined from Eq(5-30). The value of the sample statistics $[w(x)]_j$ described in Eq(5-6) using Eq(5-6), Eq(5-16)

and Eq(5-30). The cumulative probability density function described in Eq(5-4) can then be addressed.

Once the cumulative probability density function is determined, we can now use the Tail Model built in chapter 4 to get a better fit. Details of the procedure of Tail Model can be found in section 4.6 - Procedure of Tail model.

5.12 Summary

An Asymptotic Importance Sampling method has been presented based on the Tail Model built in this chapter. In this method, sampling densities are derived for a variety of practical conditions: a single point of maximum log-likelihood; points located at the interest of several failure surfaces, and bounded variables.

Theoretically, the quality and the efficiency of the method improves as the failure probability decrease, this will be demonstrated in the next chapter by applying the method to a few examples to validate the model and it will then be applied to real life applications.

Chapter 6 : Validations and Applications

6.1 Introduction

The Asymptotic Importance Sampling method has been presented based on the Tail Model built in the previous chapter. Theoretically, the quality and the efficiency of AIS Tail Model improves as the failure probability decrease. To validate the new model and to investigate the characteristics of the AIS Tail Model, a few examples will be employed in this chapter and the new model will then be used to a real-life example. Traditional methods, such as Monte Carlo Simulation (MCS), First/Second Order Reliability Method (FORM/SORM), are also used as an comparison.

Section 6.2 uses two simple example to perform the theoretical validation. Example 1 is a simple case in which two independent variables are involved with experimental distributions. The reason of this simple case is because we can get the analytical results in this example and thus provide us with the ability to proof that AIS Tail Model can give accurate results. Example 2 is also a theoretical example to investigate the effects of variable dimensions. It provides a comparison with traditional methods, Monte Carlo Simulation (MCS), First/Second Order Reliability Method (FORM/SORM), to prove that AIS Tail Model can give more accurate results with small numbers of simulations.

The new model is then used to a real life casing design example in drilling.

6.2 Theoretic Validation

Example 1:

This example is for illustrating the accuracy and parameter sensitivity of using AIS method built in the previous chapter. To give a good comparison, the example used in chapter 4 is still used for the validation.

In the simple case of two independent variables z_1 and z_2, with parameters μ and σ, the state function is as follows:

$$g(z_1, z_2) = \gamma^2 - z_1 z_2$$

the failure probability, the sensitivity of P(F) with respect to μ and σ, and the sensitivity of P(F) with respect to γ are determined respectively in this example.

The exact results can be reduced as the following,

$$P(F) = \Phi\left(-\sqrt{2}\,\frac{\ln\gamma - \mu}{\sigma}\right),$$

$$\frac{\partial P(F)}{\partial \mu} = \frac{1}{\sqrt{\pi}\sigma} e^{-\left(\frac{\ln\gamma-\mu}{\sigma}\right)^2}$$

$$\frac{\partial P(F)}{\partial \sigma} = \frac{1}{\sqrt{\pi}\sigma^2}(\ln\gamma - \mu) e^{-\left(\frac{\ln\gamma-\mu}{\sigma}\right)^2}$$

$$\frac{\partial P(F)}{\partial \gamma} = -\frac{1}{\sqrt{\pi}\sigma\gamma} e^{-\left(\frac{\ln\gamma-\mu}{\sigma}\right)^2}$$

Some of the parameters needed during the computation are as follows:

$$l(z_1, z_2) = c - 2\ln\sigma - \ln z_1 - \ln z_2 - \sum_{i=1}^{2} \frac{(\ln z_i - \mu)^2}{2\sigma^2}$$

$$PML = (\gamma, \gamma)^T, \left|\frac{\nabla l}{\nabla g}\right| = c\gamma^{-2}, \text{ where } c = \left[1 + \frac{\ln\gamma - \mu}{\sigma}\right]$$

$$\hat{x}_1 = \begin{Bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{Bmatrix} \quad k_{l2} = \frac{1}{\sqrt{2}\gamma c\sigma^2}(1 - c\sigma^2),$$

$$k_{g2} = -\frac{1}{\sqrt{2}\gamma}$$

γ=4.15148, μ=1, σ=0.150

Where $l(z_1, z_2)$ can be calculated from Equation (5-10), k_{gi} are the principal curvatures as seen in Equations (5-27), C can be calculated from (5-24).

Table 6-1 shows the comparison of the exact result and the one achieved by using AIS Tail Model built in chapter 5.

It is clear that, using small quantity of iterations, a good result can be obtained using the AIS Tail Model. The parameter sensitivity is also analysed in this example, it shows that the probability is direct proportional to μ and σ, but inverse to γ.

Table 6-1: Comparison with exact results

		$P(F)$	Error	$\partial P/\partial\mu$	Error	$\partial P/\partial\sigma$	Error	$\partial P/\partial\gamma$	Error
Exact result		3.167E-5	—	1.261E-3	—	3.569E-3	—	-3.304E-4	—
MCS(m=10,000)		3.182E-5	0.47%	1.281E-3	1.54%	3.625E-3	1.57%	-3.547E-4	7.3%
FORM		3.428E-5	8.20%	1.420E-3	12.6%	3.207E-3	10.1%	-2.953E-4	10.6%
SORM		3.250E-5	2.62%	1.316E-3	4.36%	3.372E-3	2.72%	-3.012E-4	8.84%
Tail Model	m =50	3.181E-5	0.44%	1.275E-3	1.11%	3.515E-3	1.50%	-3.010E-4	8.9%
	m =500	3.189E-5	0.69%	1.270E-3	0.71%	3.617E-3	1.34%	-3.018E-4	8.6%
	m =5000	3.170E-5	0.09%	1.264E-3	0.24%	3.552E-3	0.48%	-3.021E-4	8.5%
AIS Tail Model	m =50	3.176E-5	0.28%	1.261E-3	0.00%	3.522E-3	1.30%	-3.018E-4	8.6%
	m =500	3.189E-5	0.29%	1.268E-3	0.50%	3.614E-3	1.26%	-3.024E-4	8.4%
	m =5000	3.166E-5	0.03%	1.262E-3	0.08%	3.566E-3	0.1%	-3.021E-4	8.5%

To further investigate the accuracy and efficiency of the new model, different values of γ are chosen. Figure 6-1 shows the results from FORM, SORM, MCS, Tail Model, AIS Tail model and also the expected results with γ increases from 1.0 to 5.0. Figure 6-2 gives the absolute errors of these results. As it has been demonstrated in chapter 4, it can be seen that, as γ increases, FORM becomes more accurate. This is because the limit state function has less curvature when γ increases. AIS Tail

Model gives very much better results with small number of simulations than the Tail Model demonstrated in chapter 4. MCS provides very close results with 1,000,000 simulations.

Because FORM and SORM require only a small number of simulations they are both useful at the preliminary design stage. When more accurate results are required, the AIS tail method is better than FORM and SORM, and more efficient than MCS.

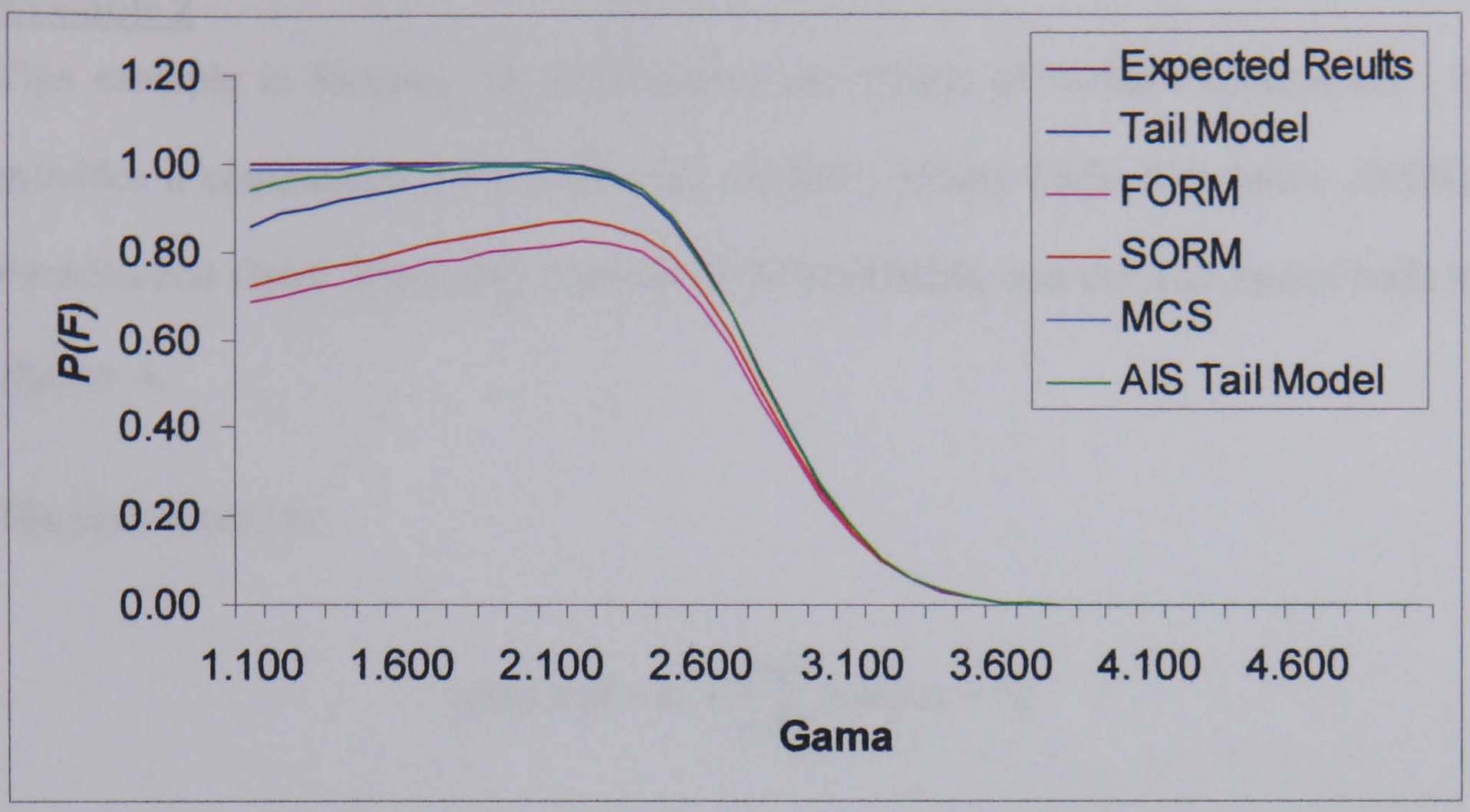


Figure 6-1: Probability results from different methods ($\mu = 1$ and $\sigma = 0.15$)

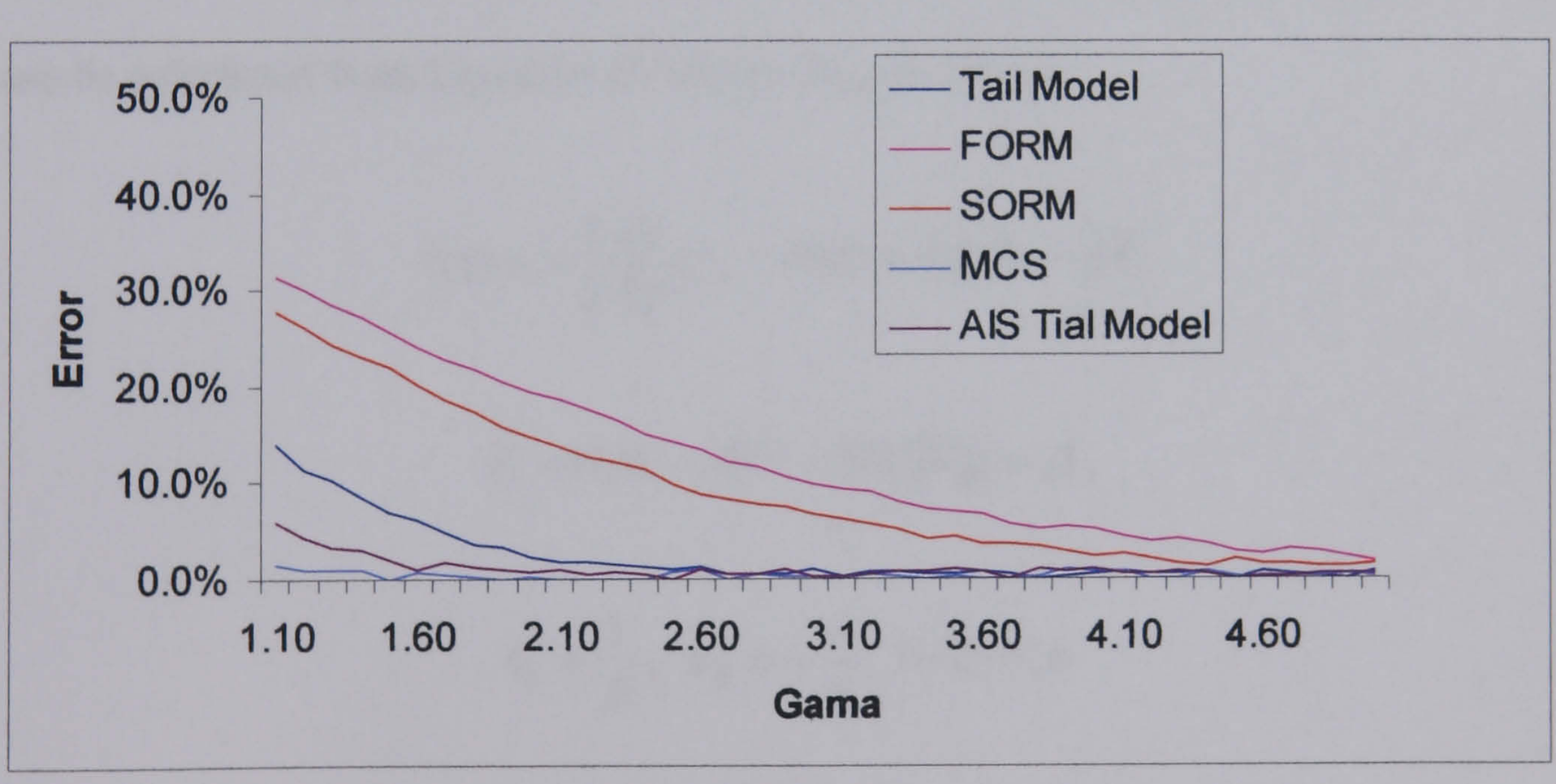


Figure 6-2: Error of the results with respect to the expected results

Example 2

This example is focusing on investigating the effects of variable dimensions. It provides a comparison with traditional methods, Monte Carlo Simulation (MCS), First/Second Order Reliability Method (FORM/SORM), and the Tail model built in chapter 4.

The state function:

$$g(z) = \beta - z_1 + \frac{a}{n}\sum_{i=2}^{n}(\cosh z_i - 1)$$

where $z_i (i = 1, \cdots, n)$ are n independent standard normally distributed random variables in interval [0,1].

Same as in Example 1, some of the parameters needed during the computation are can be calculated from Equation (5-10), (5-24), (5-27) etc.

$$l(z) = -\frac{1}{2}\sum_{i=1}^{n} z_i^2, \quad PML = (\beta, 0, \cdots, 0)^T$$

$$\hat{x}_1 = (1, 0, \cdots, 0)^T, \quad |\nabla l| / |\nabla g| = \beta,$$

$$k_{li} = \frac{1}{\beta}, \quad k_{gi} = -\frac{\alpha}{n}, \quad i = 2, \cdots, n$$

Same as in Chapter 4, the Monte Carlo results after 1,000,000 iteration is introduced here as an "accurate" result to calculate the error. Table 6-2, Table 6-3, Table 6-4 and Table 6-5 show the results using different simulation methods under different α and β to investigate the effects of the variable dimension.

It can be seen that FORM, in contrast to FORM, SORM and the Tail Model built in Chapter 4, AIS Tail model provides very accurate results even in the case of high dimensionality with small number of simulations. The error increases when the dimensionality becomes higher, the accuracy can be increased in AIS Tail Model by enlarging the number of iteration steps.

Table 6-2: Comparison of different methods under different dimension (β=5, α=1)

		$n=2$	Error	$n=10$	Error	$n=50$	Error
MCS (m=1,000,000)		1.5120E-07	-	5.6500E-08	-	1.2500E-08	-
FORM		3.0100E-07	99.07%	3.0100E-07	432.74%	3.0100E-07	2308.00%
SORM		1.6045E-07	6.12%	7.0160E-08	24.18%	2.8370E-08	126.96%
Tail Model	m=50	1.6065E-07	6.25%	6.4037E-08	13.34%	1.7074E-08	36.59%
	m=500	1.5719E-07	3.96%	6.1432E-08	8.73%	1.5640E-08	25.12%
	m=5000	1.5344E-07	1.48%	5.9478E-08	5.27%	1.5008E-08	20.06%
AIS Tail Model	m=50	1.5366E-07	1.63%	5.9235E-08	4.84%	1.5270E-08	22.16%
	m=500	1.5264E-07	0.95%	5.8042E-08	2.73%	1.4885E-08	19.08%
	m=5000	1.5199E-07	0.52%	5.7636E-08	2.01%	1.4699E-08	17.59%

Table 6-3: Comparison of different methods under different dimension (β=3.5, α=1)

		n=2	Error	n=10	Error	n=50	Error
MCS (m=1,000,000)		1.332E-4	-	4.229E-5	-	2.913E-5	-
FORM		2.326E-4	74%	2.326E-4	400%	2.326E-4	200%
SORM		1.403E-4	5.3%	6.028E-5	42%	4.434E-5	52%
Tail Model	m=50	1.410E-4	5.8%	4.750E-5	12.3%	1.965E-5	32.5%
	m=500	1.290E-4	3.1%	4.602E-5	8.8%	2.158E-5	25.9%
	m=5000	1.315E-4	1.3%	4.513E-5	6.7%	2.250E-5	22.7%
AIS Tail Model	m=50	1.350E-4	1.4%	4.050E-5	4.2%	2.273E-5	22%
	m=500	1.324E-4	0.6%	4.119E-5	2.6%	2.290E-5	21%
	m=5000	1.325E-4	0.5%	4.163E-5	1.6%	2.293E-5	21%

Table 6-4: Comparison of different methods under different dimension (β=3, α=-0.1)

		$n=2$	Error	$n=10$	Error	$n=50$	Error
MCS (m=1,000,000)		1.5200E-03	-	1.6496E-03	-	1.5200E-03	-
FORM		2.6700E-03	75.66%	2.6700E-03	61.86%	2.6700E-03	75.66%
SORM		1.6200E-03	6.58%	2.2800E-03	38.22%	3.2224E-03	112.00%
Tail Model	m=50	1.6086E-03	5.83%	1.8317E-03	11.04%	1.9863E-03	30.68%
	m=500	1.5705E-03	3.32%	1.7755E-03	7.63%	1.9178E-03	26.17%
	m=5000	1.5408E-03	1.37%	1.7159E-03	4.02%	1.8061E-03	18.82%
AIS Tail Model	m=50	1.5416E-03	1.42%	1.7197E-03	4.25%	1.8114E-03	19.17%
	m=500	1.5331E-03	0.86%	1.6993E-03	3.01%	1.8032E-03	18.63%
	m=5000	1.5262E-03	0.41%	1.6813E-03	1.92%	1.7703E-03	16.47%

Table 6-5: Comparison of different methods under different dimension (β=3, α=0)

		$n=2$	Error	$n=10$	Error	$n=50$	Error
MCS (m=1,000,000)		1.3604E-01	-	1.3604E-01	-	1.3604E-01	-
FORM		1.3489E-01	0.84%	1.3489E-01	0.84%	1.3489E-01	0.84%
SORM		1.3513E-01	0.67%	1.3513E-01	0.67%	1.3513E-01	0.67%
Tail Model	m=50	1.3539E-01	0.48%	1.3539E-01	0.48%	1.3539E-01	0.48%
	m=500	1.3652E-01	0.35%	1.3652E-01	0.35%	1.3652E-01	0.35%
	m=5000	1.3629E-01	0.18%	1.3629E-01	0.18%	1.3629E-01	0.18%
AIS Tail Model	m=50	1.3624E-01	0.15%	1.3624E-01	0.15%	1.3624E-01	0.15%
	m=500	1.3616E-01	0.09%	1.3616E-01	0.09%	1.3616E-01	0.09%
	m=5000	1.3610E-01	0.04%	1.3610E-01	0.04%	1.3610E-01	0.04%

6.3 Real life Application

The validation has been done in the former section. It shows that the new method is more accurate than the existing simulation methods. In this part, the method is used into the real life casing design in oil and gas industry.

While drilling oil and gas wells, the most frequent failure mode is casing collapse, which is assumed to occur once the external pressure exceeds the ultimate collapse strength. To estimate the casing failure probability under collapse load, the Tamano equations for ultimate collapse strength are summarised below.

The Tamano equations for ultimate collapse strength are summarised below; a fuller treatment can be found in Tamano et al. 1983.

$$Z = \text{Res} - \text{Load}$$

$$\text{Res} = \frac{P_e + P_p}{2} - \sqrt{\frac{(P_e - P_p)^2}{4} + P_e * P_p * H}$$

Where,

$$P_e = \frac{2E}{1-\mu^2} * \frac{1}{\frac{D}{t} * (\frac{D}{t} - 1)^2}$$

$$P_p = 2 * Y * (\frac{D}{t} - 1)(\frac{t}{D})^2 * (1 + \frac{1.47}{\frac{D}{t} - 1})$$

The effect of eccentricity, ovality is modelled by a strength decrement factor H, whose coefficients were obtained from elasto-plastic finite element analysis (Tamano et al. 1983).

$$H = 0.0808 * OV + 0.00114 * EC + 0.1412 * \frac{\sigma_y}{\mu f_y}$$

Where variables T, D are casing average thickness and average outside diameter; E, Y, μ and RS are Young's modulus, yield stress, Poison's ratio and residual stress individually; EC and OV which are eccentricity and ovality are considered in the equation. The mean and covariance of each variable are shown in Table 6-6.

Table 6-6: Basic variables

Variable	Nominal Value	Mean Value	COV
Yield Stress	Y=379,211,800 Pa	1.09	0.03
Casing Avg thickness	T=12.7 mm	1.0	0.025
Casing Avg OD	D=508 mm	1.005	0.001
Young's Modulus	E=3.0E7	1.0	0.035
Poison's Ratio	μ=0.3	1.0	0.025
Eccentricity	EC=7.4 %	1.0	0.51
Ovality	OV=0.18 %	1.0	0.54
Residual Stress	RS=-1447.90 Pa	1.0	0.20
Mud Weight Prev String	MWCSD= 226,206 kg/m^3	1.0	0.06
Mud Weight Next String	MWTD= 271,447kg/m^3	1.0	0.06
Lost Circ Depth	Lcdepth=127m		
Pore Pressure	PoreP= 13,100,044 Pa	1.0	0.06

Table 6-7 gives the distribution type of the variables involved in the design equation. Some variables, which enough data set can be found, were treated as random ones in the simulation procedure. The others were still considered as it shows in DEA64. Appendix E lists the PDF data for variables.

Table 6-7: The distribution type of each variables in the design equation

	Variable	Distribution
D	Outside diameter	Random
T	Wall thickness	Random
Ec	Eccentricity	Random
Ov	Ovality	Random
E	Young's modulus	Gaussian
μ	Poisson's ratio	Gaussian
σ_y	Yield stress	Gaussian

The following simulations are to investigate the characteristics of the equation under different load case, meanwhile, the comparison between different method is also presented.

6.3.1 Investigation of Lost Circulation

According to DEA 64, the lost circulation load case is as follows:

$$L = \rho_o g(d - \frac{pp_d}{\rho_i g})$$

where pp_d is pore pressure at lost circulation depth and d is lost circulation depth,

ρ_o and ρ_i are fluid density outside of casing and inside of casing respectively. Table 6-8 gives the distribution type of variables involved in the lost circulation load case.

Table 6-8: Distribution type of variables in lost circulation load case

	Variable	Distribution
pp_d	Pore pressure lost circulation depth	Log-Normal
d	Lost circulation depth	Log-Normal
ρ_o	Fluid density outside casing	Gaussian
ρ_i	Fluid density inside casing	Gaussian
g	Gravity Acceleration	
μf_y	Mean value of yield stress	

A set of casing with different D/t was analysed by the AIS method (where m=50), the corresponding failure probability is shown in Figure 6-3. For comparison, the MCS result is also plotted in the same graph.

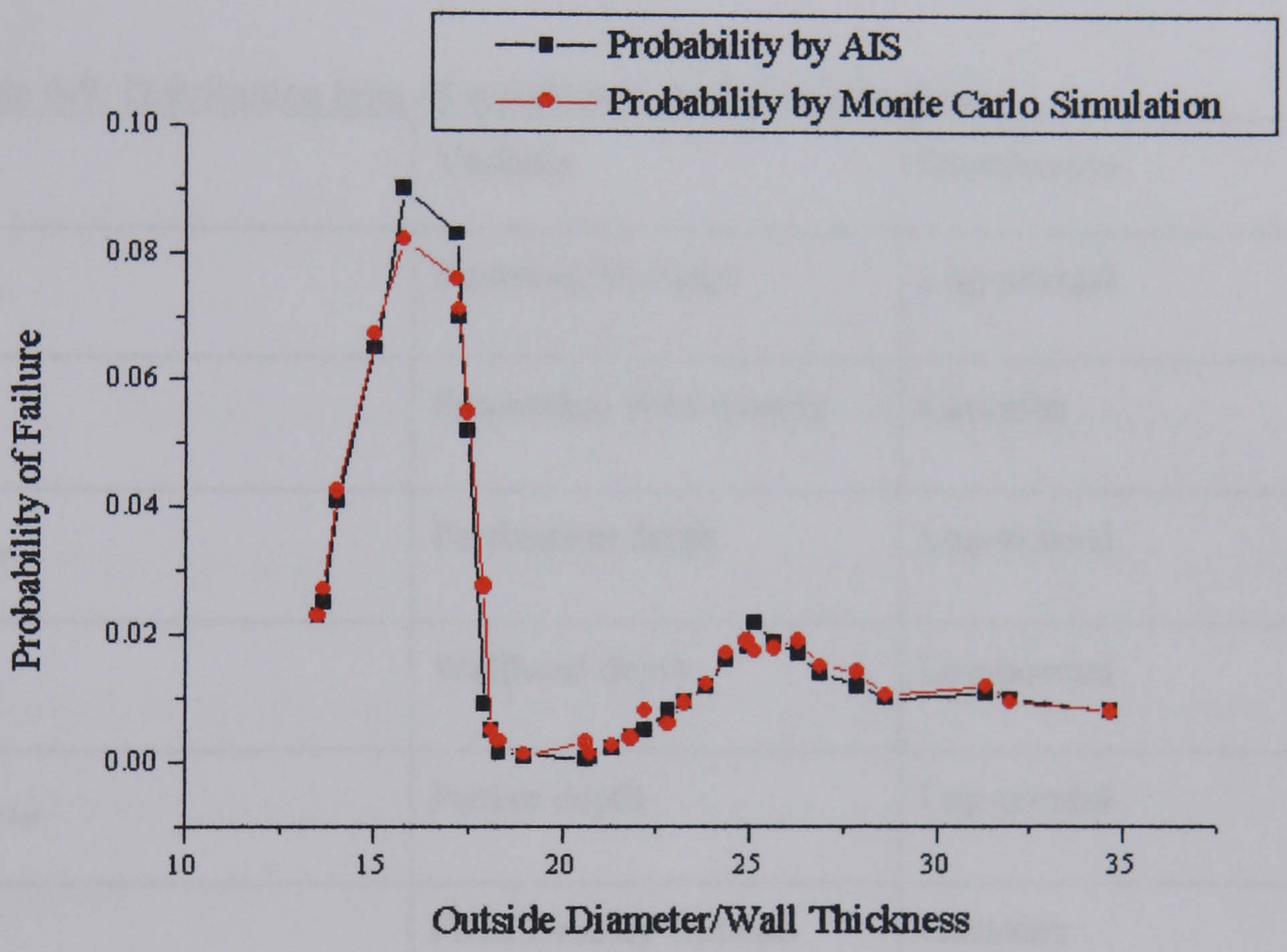


Figure 6-3: Failure probability to D/t of Tamano Equation under lost circulation.

6.3.2 Analysis for tubing leak load case

Kill Weight Probability function:

$$L = [p_{res} - \rho_{pf} g(d_{perfs} - d_{wh}) + \rho_i g(d_{packer} - d_{wh})] - \rho_0 g d_{packer}$$

Light Weight probability function

$$L = [p_{res} - \rho_{pf} g(d_{perfs} - d_{wh})] - \rho_0 g d_{wh}$$

The distribution information lists in Table 6-9. A set of casing with different D/t was analysed by the AIS method (where m=50), the corresponding failure probability is shown in Figure 6-4. Results of SORM and MCS are also plotted for comparison.

Table 6-9: Distribution type of variables in tubing leak load case

	Variable	Distribution
p_{res}	Reservoir Pressure	Log-normal
ρ_{pf}	Production fluid density	Gaussian
d_{perfs}	Perforation depth	Log-normal
d_{wh}	Wellhead depth	Log-normal
d_{packer}	Packer depth	Log-normal
ρ_0	Fluid Density Outside Casing	Gaussian
g	Acceleration due to gravity	

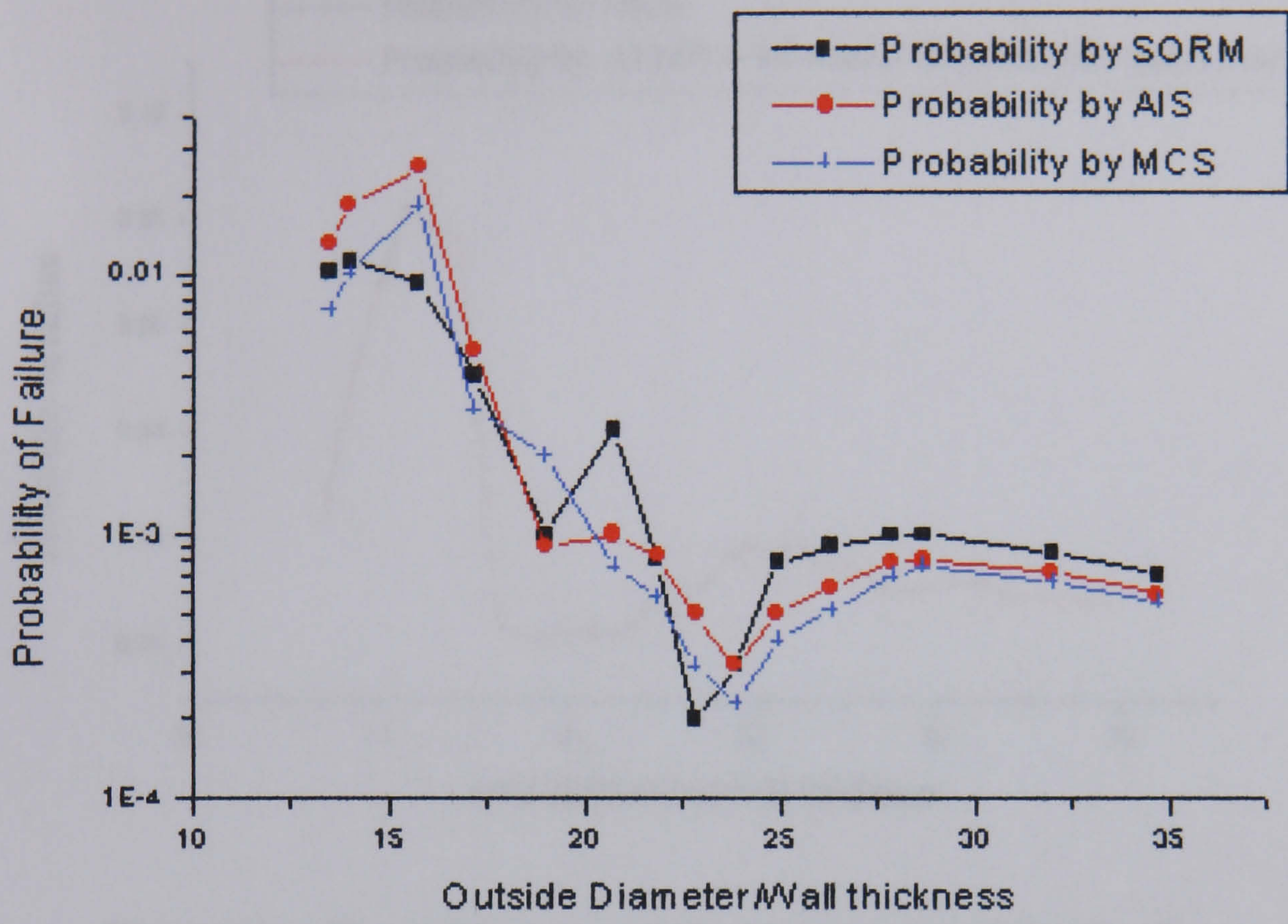


Figure 6-4: Casing failure probability under tubing leak load case

6.3.3 Investigating of the effects of different variables to the results

To investigate the advantages and accuracy of AIS method, this part introduces only a few variables as random and the rest of them are still treated as deterministic or distributed as stated in DEA64,

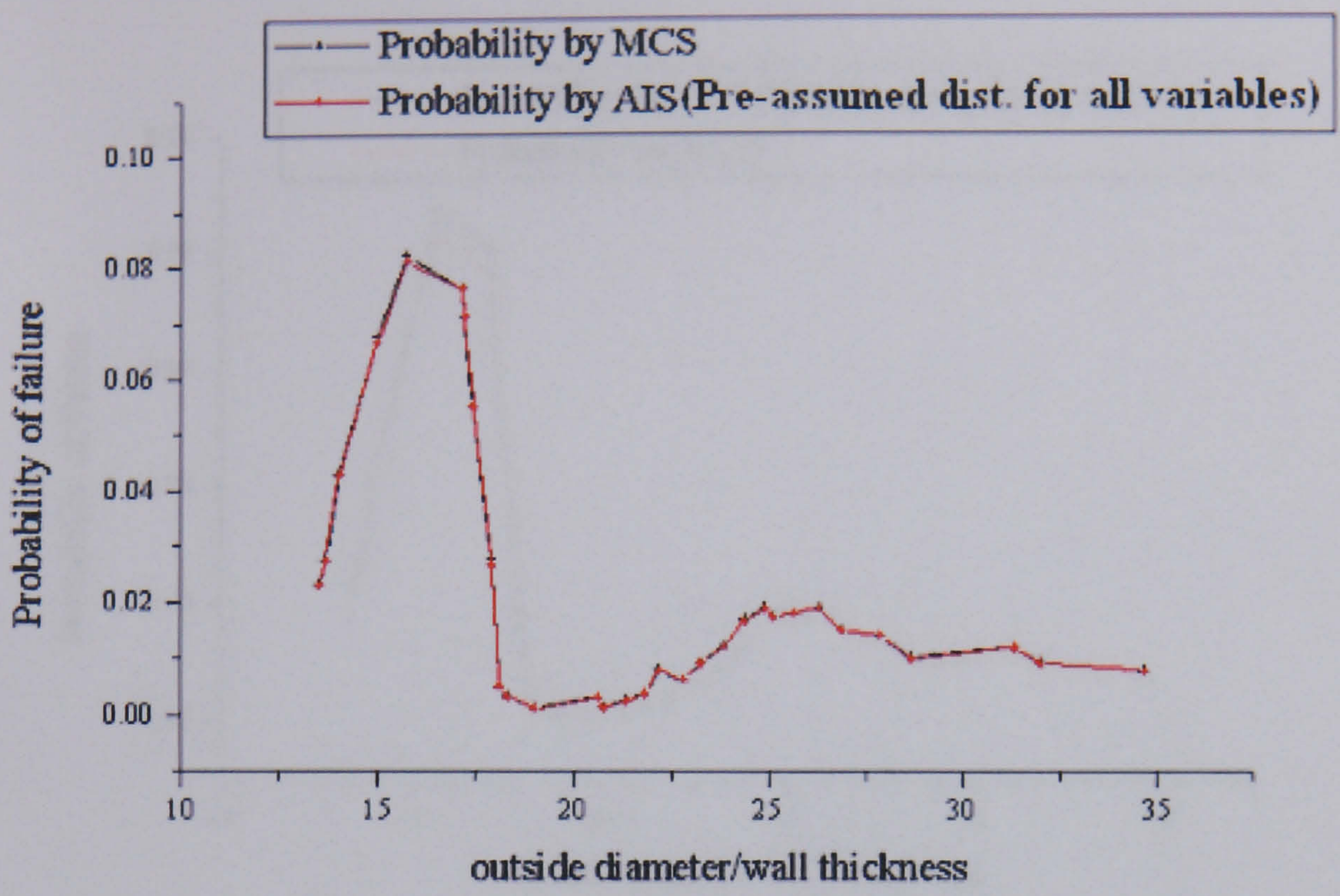


Figure 6-5: All variables are treated with pre-assumed distributions

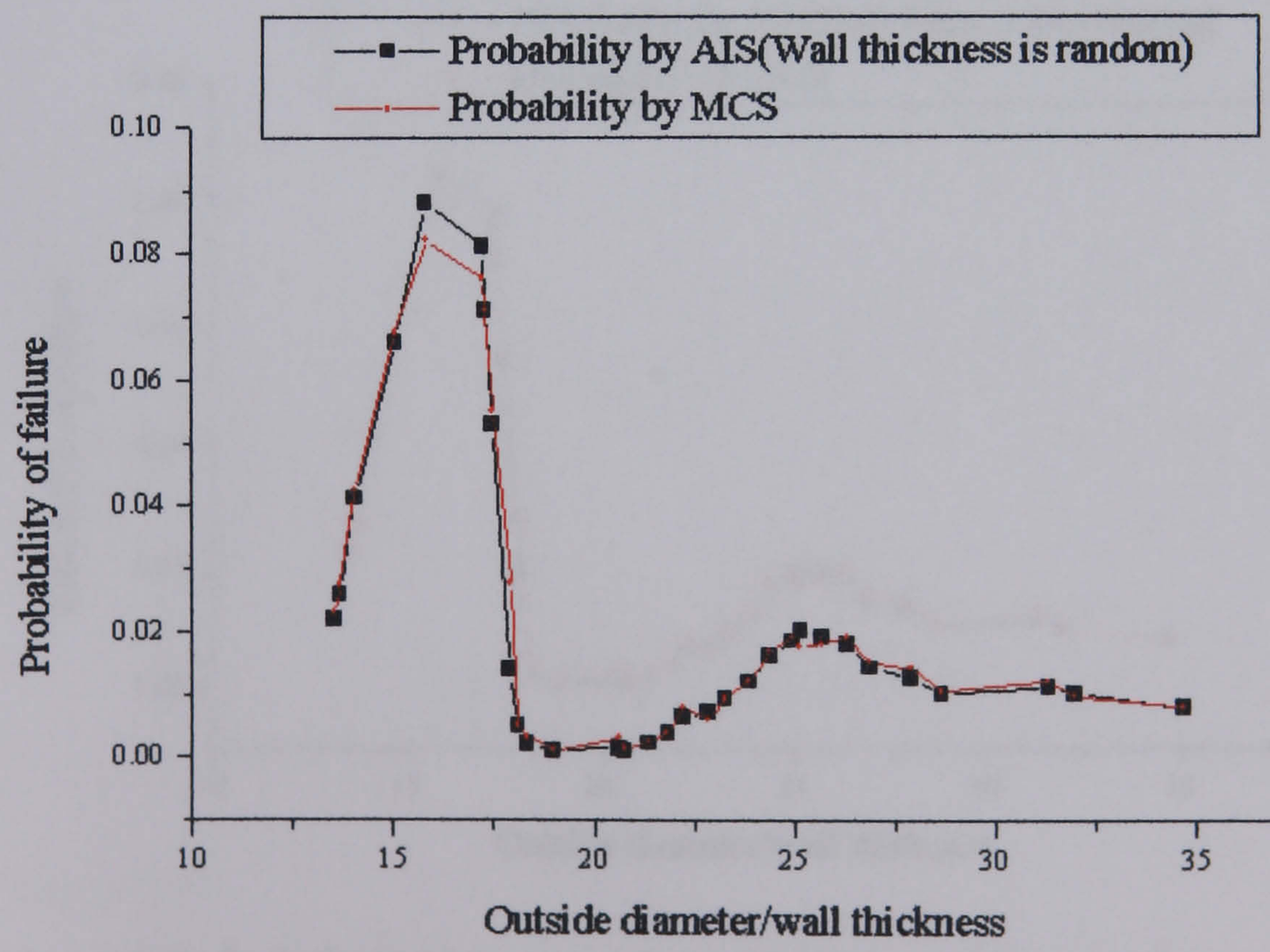


Figure 6-6: Only wall thickness is treated as random

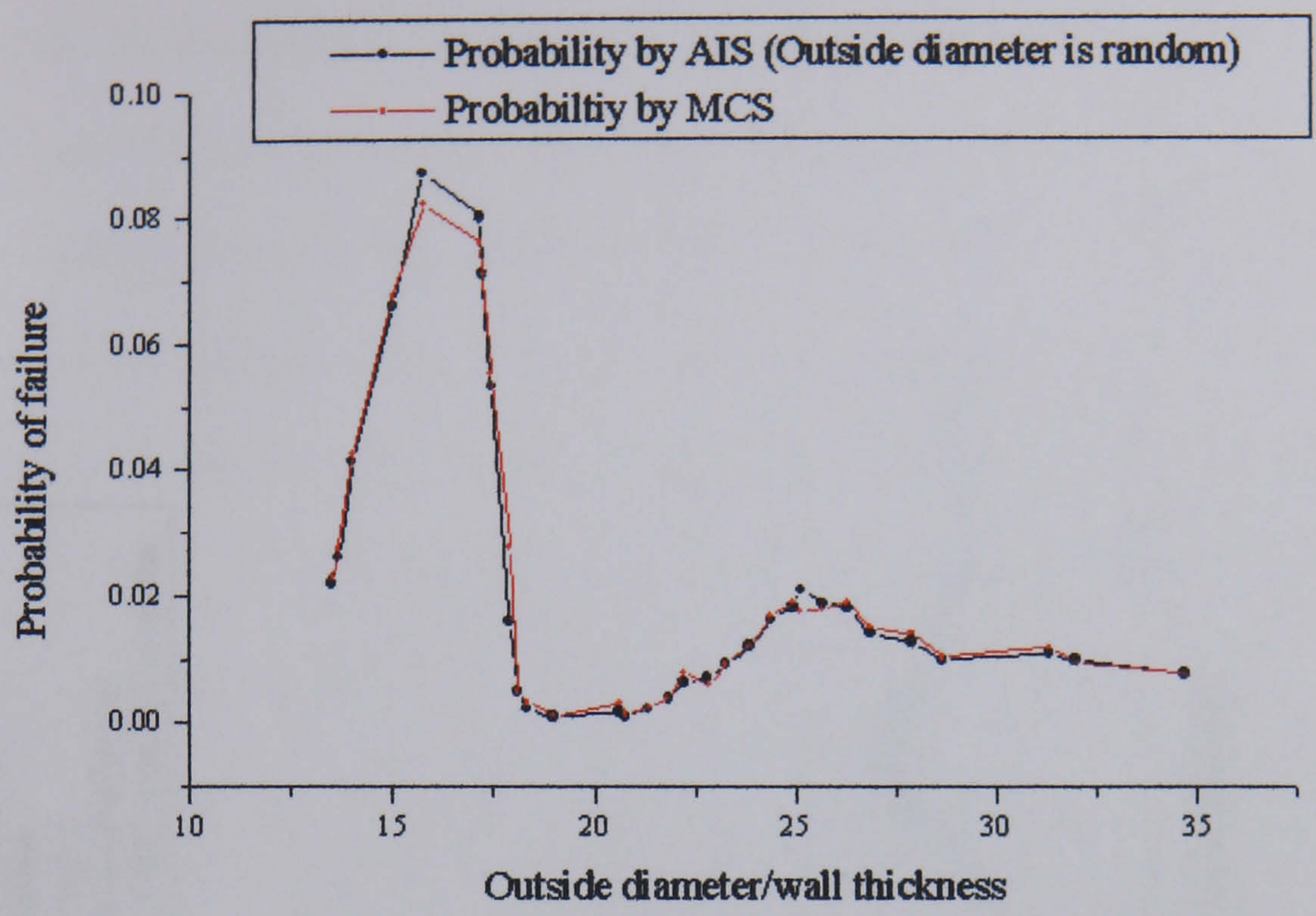


Figure 6-7: Only outside diameter is treated as random

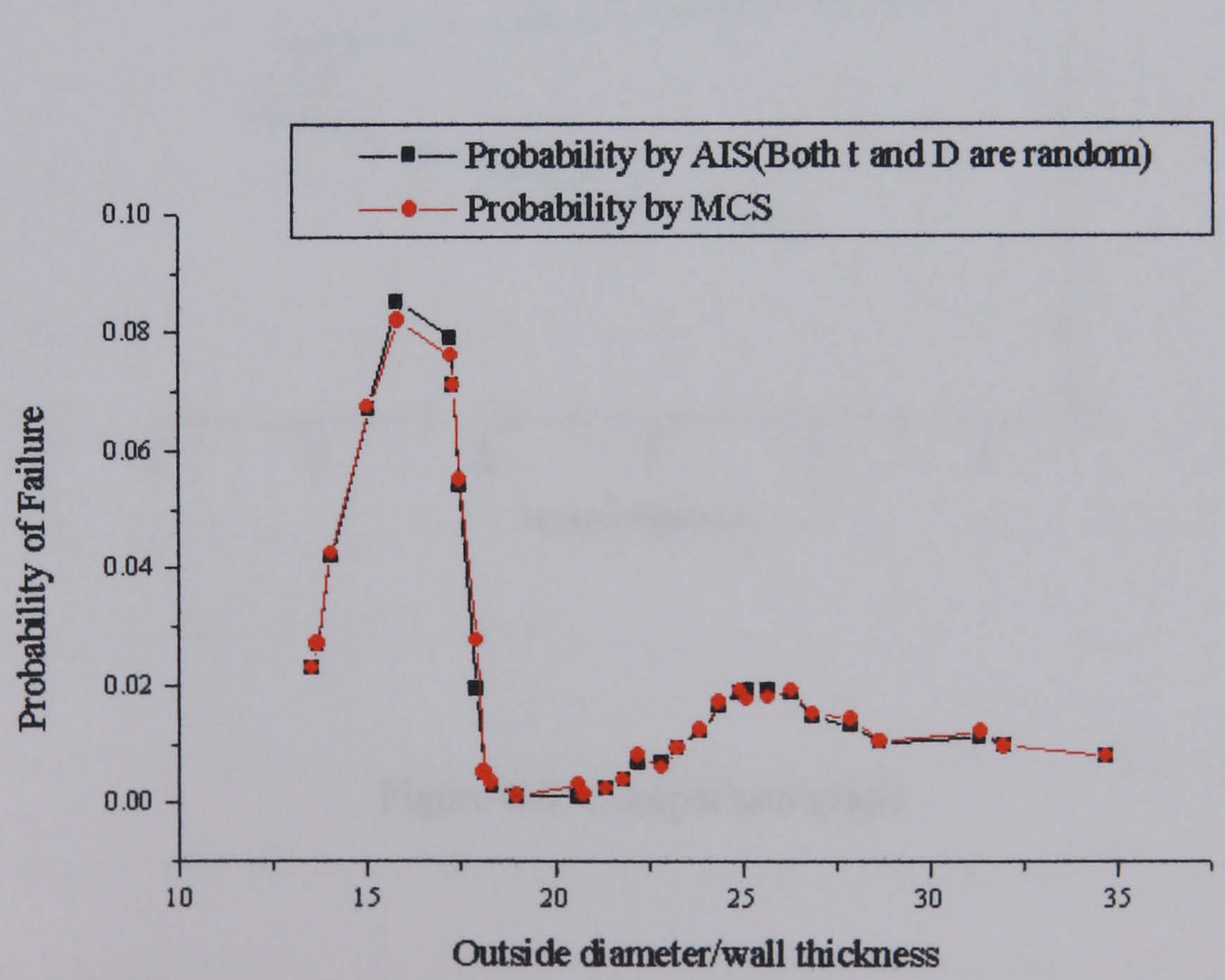


Figure 6-8: Both Wall thickness and outside diameter are treated as random:

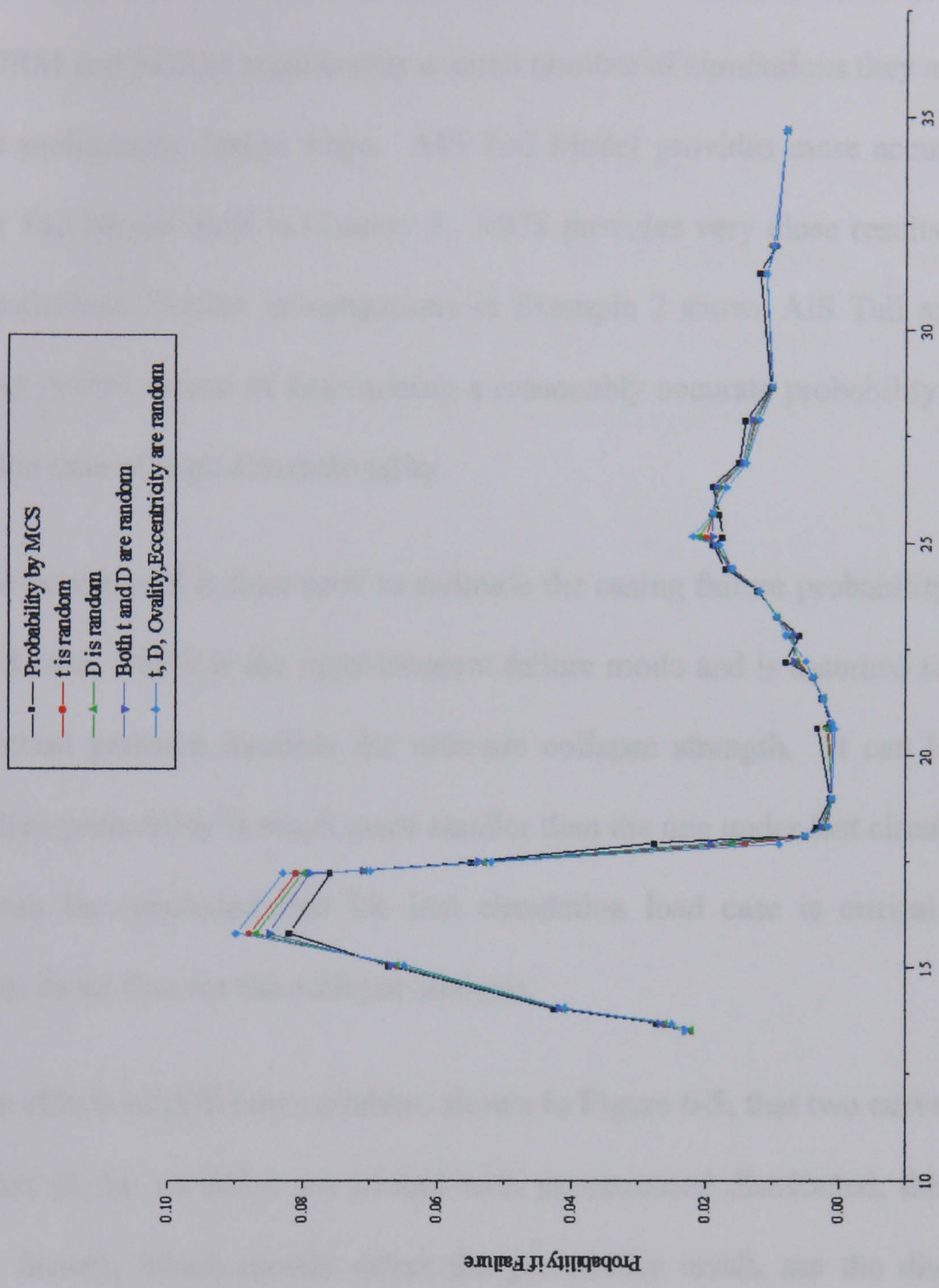


Figure 6-9: Comparison graph

6.4 Summary

The same examples as in Chapter 4 are used in this chapter to validate the model. Example 1 has shown the comparison of the exact result with those returned by

FORM, SORM, MCS, Tail Model and AIS Tail Model. As γ increases, because the limit state function has less curvature, FORM becomes more accurate. Because FORM and SORM require only a small number of simulations they are both useful at the preliminary design stage. AIS Tail Model provides more accurate results than the Tail Model built in Chapter 4. MCS provides very close results with 1,000,000 simulations. Further investigations in Example 2 shows AIS Tail model provides a more robust means of determining a reasonably accurate probability of failure, even in the case of high dimensionality.

The new model is then used to estimate the casing failure probability under collapse load case, which is the most frequent failure mode and is assumed to occur once the external pressure exceeds the ultimate collapse strength. It can be seen that the failure probability is much more smaller than the one under lost circulation load case. It can be concluded that the lost circulation load case is critical and should be considered first for the collapse analysis.

The effects of different variables, shown in Figure 6-5, that two curves are very close when all the variables are treated with pre-assumed distributed, this illustrates that the factors, which mostly effect the probability result, are the distribution of the variables rather than the method used.

When the wall thickness and/or the outside diameter were treated as random variables and fitted by the method developed, as shown in Figure 6-6, Figure 6-7 and Figure 6-8 respectively, the results are different especially when D/t is between 15~18, where the curve is very steep. This shows that AIS Tail model is not well

suited to large probability of failure which tail area does less contribution to the results.

Because the AIS method offers a theoretical basis for getting a suitable distribution type of all variables with limited data, it can therefore be concluded that AIS Tail Model offers a great advantages in practical applications. That is to say, the distribution type is not essential when using AIS Tail Model. With limited data, AIS Tail Model can give more accurate results in comparison with traditional methods.

Chapter 7 : Discussion

The present research started with a review of the existing casing design methodology in the oil and gas industry.

As the main method used in the traditional casing design, Working Stress Design uses design factors, which are almost entirely empirical, to account for the uncertainties of the design variables, and its minimum expected strength is designed to exceed maximum possible loads. Because it is always based on the worst case of wells for which the failure probability is the highest, the selected casing string is normally over-sized for the less severe cases such as development wells. The uses of design factors is normally used for all steel types and grades, for all strings and all load types, which causes uneven risk as discussed in chapter 1.

Because Quantitative Risk Assessment provides the methodology for the optimisation of a design, while quantifying the risk of the design for a particular scenario and can give the designer a satisfactory and quantified answer, the existing QRA methods were then investigated.

First/Second Order Reliability Method (FORM/SORM) and Monte Carlo Simulation (MCS) are at present very popular in Quantitative Risk Assessment. The advantages of these methods are that they give a simple procedure and its algorithms are easy to be implemented. A major disadvantage, however, is that it is very computationally intensive when the simulated event is of very low probability. Meanwhile, in QRA simulation procedure, MCS and FORM/SORM focus on the theoretical distribution in the central part of the observed data, rather than on the tail.

The aim of the new methodology is to build up a mathematical that simulates the data of the reality better than the existing methods and can quantify the risk of failure more precisely.

The investigation of existing Quantitative Risk Assessment methods shows that, in the casing design of oil and gas industry, the tail area is most effective to the probability result. That is to say, whether or not the results are accurate enough critically depends on how good the mathematical model can simulate the tail area. Therefore, a way of estimating the tail area of a distribution is presented in the research.

The key idea of the Tail Model is based on introducing the tail heaviness index (THI) into the generalised Pareto Distribution which is a two-parameter distribution that contains the uniform, exponential, and Pareto distribution as special case. THI describes how heavy any portion of a probability density tail is with respect to the pure exponential tail whose THI equals zero, and it is negative for lighter than exponential tails (sub-exponential) and positive for heavier than exponential tails (super-exponential).

Studies on comparison of traditional QRA simulation methods with the Tail Model shows that, for the examples considered, FORM provides a very inaccurate result even if only two variables involved. The error is getting larger when the dimensionality increases. The accuracy of SORM estimate is acceptable for low dimensionality.

Although in certain special cases FORM and SORM can be accurate and efficient. Tail model provides a more robust means of determining a reasonably accurate probability of failure, even in the case of high dimensionality. The result that

provided by the Tail model is very close to the one provided by MCS when only two variables involved in the simulation. Similar to FORM and SORM, the error of Tail Model increases rapidly as the dimension of variable space becomes higher. Obviously, this new methodology can not satisfy the use of casing design because, in casing design, a lot of variables will appear in the design equation, such as outside diameter, wall thickness, ovality, eccentricity, etc. It is normally more than 10 variables. That is to say, if the probability model has a larger numbers of variables, the Tail Model will not be able to simulate the reality very well.

To overcome the disadvantages of the Tail Model, Asymptotic Important Sampling (AIS) method was introduced. AIS techniques are an efficient technique for multivariate integration especially when the probability is small. The key idea of AIS is based on the careful selection of a sampling density for subsequent use in an importance sampling scheme. The selection is based on theoretical consideration of the structure of the integration domain in the original variable space. Due to the asymptotic properties of the sampling densities, this technique makes the Tail model increasingly efficient in terms of obtaining a given level of accuracy, as the failure probability becomes smaller.

Chapter 8 : Conclusion

- The traditional casing design method and QRA methods were investigated. QRA methods have been applied in the design of casing and it has been proven that QRA can give the user a quantitative answer.

- A new QRA methodology, AIS Tail Model, was built and used for the real life casing design in the oil and gas industry.

- Studies shows in contrast to FORM and SORM, AIS Tail model provides a more robust means of determining a reasonably accurate probability of failure, even in the case of high dimensionality for the number of examples considered.

- The AIS Tail Model needs less iterations to achieve an accurate result, and is especially efficient when large numbers of variables are involved in the design equation.

- The new methodology uses random variables and does not require the choice of a distribution type, while traditional design methods use a pre-determined distribution type. It avoids the rather arbitrary selection of a cut-off point between the central part and the tail part of a data set. It is consistent with simple risk measures and yields an estimate of the associated uncertainty of the model.

- The AIS Tail Model is an approximation which is good in an asymptotic sense, its quality depends on the overall structure of the LSS and the loglikelihood function. But for small $P(F)$, the new model is guaranteed to give more accurate results than the traditional methods.

- QRACD (Quantitative Risk Assessment in Casing Design) has been developed for ease of use of the new method in casing design. The existing methods were built into the software also; and, QRACD provides user a friendly user interface for the new and existing methods.

- Theoretically, AIS Tail Model can be used for other statistical analyse. To investigate the generalisation of the new method it was used in a real life example of share price prediction in Appendix G.

Chapter 9 : Future Work

Quantitative Risk Assessment has been introduced into failure probability analysis of drilling casing design in the oil and gas industry. From the investigation in the former chapters, a few new research topics have been opened up.

How to get an accurate result is the most critical factor in the design procedure. From the examples discussed in Chapter 6, it has been found that the standard error is becoming larger when the dimension of variable space increases. Especially in example 2 of section 6.1, it shows that when the dimensionality of the variable space $n = 50$, the standard errors of both existing and AIS model are too large to be acceptable although AIS can give a more accurate result. Although the new method can achieve enough accuracy for casing design because there are normally around 10 to 20 variables involved in the casing design equation, theoretically, more work can be done to prove the accuracy so that this method can be used in other research areas.

Collapse and burst analyses have been implemented in the present research. However, to estimate the failure probability properly to ensure the safety, another failure mode, failure of axial tension, has to be done as well.

All the variables involved in the load are still treated with deterministic distribution because not enough data can be found. However, it is obvious that load is the most uncertainty factor that effects the failure. Therefore, more research on determining and quantify the uncertainty of load should be done.

To give a guideline of the failure probability, analyses have to be done in a large range D/t and large range of grades, while the present research just based on L-80. It would be very useful if a failure probability of all the grades can be given so that user can have a larger range of choice, on both D/t and grades.

Furthermore, as the casing program of most oil and gas wells represents the greatest single item of expense in well cost. It can be as much as 18% of the completed well cost. Therefore, even a small reduction in casing cost can save a considerable amount of money. The present research just aims at optimising the selection of outside diameter of the casing. To minimise the cost, consideration of the possibility of several combinations of different casing grade, weight and thread selection appears to be increasingly important.

As it states in Appendix F, the software still needs lots of further development, thesis includes:

- Build-in more simulation methods, FORM/SORM, Latin Hyper-cube Simulation, etc., and provide user the guideline.

- The program should allow user to add new data to enlarge the database so that the software can be more close to an expert system of casing design.

- Currently, the software just provides the unique grade analysis (default as L-80), the capability of analysing all grades of casing should be put into the software.

- If the new mini-cost casing design procedure, as stated above, can be developed and built into the software, it will provide user a completed casing design, on selection of outside diameter, grades and the casing string.

Chapter 10 : References

[1] Brand, W.S. Whitney and D.B. Lewis, "Load and Resistance Factor Design Case Histories " , 27th Annual OTC in Houston, Texas, U.S.A, 1-4 May, 1995.

[2] Maes, K.C. Gulati, D.L.Mckenna, P.R. Brand, D.B. Lewis, R.C. Johnson, "Reliability Based Casing Design " , Journal of Energy Resources Technology, 1993.

[3] Maes, K.C. Gulati, D.L. Mckenna, P.R. Brand, D.B. Lewis, R.C. Johnson, "Reliability-Based Casing Design: Principle and Approach ".

[4] Adams, S.H.L. Parfitt, T.B. Reeves, J.L. Thorogood, "Casing System Risk Analysis Using Structural Reliability " , SPE/IDAC Drilling Conference, 1993.

[5] Marc Nunn, "Quantitative Risk Assessment in Well Design " , British Gas Research & Technology.

[6] D.B.Lewis, P.R. Brand, W.S. Whitney, "Load and Resistance Factor Design for Oil Country Tubular Goods " , 27th Annual OTC in Houston, Texas, U.S.A, 1-4 May 1995.

[7] DEA(E)-64: "The Use of QRA in Casing/Tubing Design".

[8] Gulati, D.L. McKenna, M.A. Maes, "Reliability-Based design and Application of Drilling Tubular " , 26th Annual OTC in Houston, Texas, U.S.A., 2-5 May 1994.

[9] Averill M.Law, W.David Kelton, "Simulation Modeling and Analysis(Second Edition) " , McGraw-Hill International Editions, 1991.

[10] K. Breitung, "Probability Approximations by Log-likelihood Maximisation", J. Eng. Mech., 117(3) 1997

[11] A. Nethercot, "Limit States Design of Structural Steelwork " , Second Edition, Chapman & Hall, London, UK.

[12] Neal J. Adams (1985), "Drilling Engineering: A Complete Well Planning Approach" , Penn Well Publishing Company, 1421 South Sheridan Road, P. O. Box 1260, Tulsa, Oklahoma 74101.

[13] MacGinley and T. C. Ang, "Structural Steelwork: Design to Limit State Theory", Second edition, Butterworth-Heinemann Ltd, Linacre House, Jordan Hill, Oxford OX2 8DP.

[14] Motzkin, T. S. 1956. "The assignment problem", Proc. Symp. Appl. Math. VI. Numerical Analysis. p109-125. American Mathematical Society. New York, N.Y.

[15] Tompkins, C. 1956. "Permutation problems". Proc. Symp. Appl. Math. VI. Numerical Analysis. p195-211. American Mathematical Society. New York, N.Y.

[16] Beardwood, J. E., J. H. Halton & J. M. Hammersley. 1959. "The shortest path through many points", Proc. Cambridge Phil. Soc. 55: 299-327.

[17] Hartree, D. R. 1958. "Numerical Analysis". : 84-85. 2nd ed. Oxford Univ. Press. London, England.

[18] Fisher, R. A. 1947, "The Design of Experiments. 4th ed". Oliver & Boyd. Edinburgh, Scotland.

[19] Allen, C. D. 1959. "A method for the reduction of empirical multivariable functions, Computer J. 1: 190--200.

[20] Hammersley, J. M. & K. W. Mortox. 1956. "A new Monte Carlo technique: antithetic varieties", Proc. Cambridge Phil. Soc. 52: 449-475.

[21] Hammersley, J. M. & J. G. Mauldox. 1956. "General principles of antithetic variaties", Proc. Cambridge Phil. Soc. 52: 476-481.

[22] Handscomb, D. C. 1958. "Proof of the antithetic-varieties theorem for n>2 Proc", Cambridge Phil. Soc. 54: 300-301.

[23] Halton, J. H. & D. C. Handscomb. 1957. "A method for increasing the efficiency of Monte Carlo integration", J. Assoc. Comp. Mach. 4: 329-340.

[24] Morton, K. W. 1957. "A generalisation of the antithetic variety technique for evaluating integrals", J. Math and Phys. 36: 289-293.

[25] Lowan, A. N. 1944. "Tables of Lagrangian interpolation coefficients". Mathematical Tables Project. :xxxii-xxxiii. Columbia Univ. Press, New York, X. Y.

[26] Curtiss, J. H. 1949. "Sampling methods applied to differential and difference equations", Proc. Seminar on Scientific Computation. I. B. M. Corp. New York.

[27] Taussky, O. & J. Todd. 1956. "Generation and testing of pseudo-random numbers", In Symposium on Monte Carlo Methods. : 15-28. Wiley. New York, N.Y.

[28] Richtmyer, R. D. 1958. "A non-random sampling method, based on congruence, for Monte Carlo problems", A. E. C. Research and Development Report NYO-8674 Physics. A. E. C. Computing and Applied Mathematics Center, New York Univ. New York, N. Y.

[29] Roth, K. F. 1954. "On irregularities of distribution". Mathematics. 1: 73-79.

[30] Van Der Corput, J. C. 1935. "Monte Carlo Methods". Proc. Koninkl. Ned. Akad. Wetenschap. 38: 813-821.

[31] Van Aardenne-Ehrenfest, T. 1945. "Proof of the impossibility of a just distribution of an finite sequence of points over an interval". Proc. Koninkl. Ned. Akad. Wetenschap. 48:266-271.

[32] Van Aardenne-Ehrenfest, T. 1949. "On the impossibility of a just distribution of a just distribution". Proc. Koninkl. Ned. Akad. Wetenschap. 52:734-739.

[33] COLES, W. J. 1957. "On a theorem of van der Corput on uniform distribution". Proc. Cambridge Phil. Soc. 53: 781-789.

[34] DE BRUIJN, N. G. & P. ERDOS. 1949. "Sequences of points on a circle". Proc. Koninkl. Ned. Akad. Wetenschap. 52: 46-49.

[35] MAHLER, K. 1957. "On the fractional parts of the powers of a rational number", II. Mathematika. 4: 122-124.

[36] J.M. Hammersley and D.C. Handscomb, "Monte Carlo Methods", Methuen, London, 1964

[37] TROTTER, H. F. & J. W. Tukey. 1956. "Conditional Monte Carlo for normal samples. In Symposium on Monte Carlo Methods". : 64-79. Wiley. New York, N. Y.

[38] HAMMERSLEY, J. M. 1956. "Conditional Monte Carlo". J. Assoc. Comp. Mac 3: 73-76. .

[39] WENDEL, J. G. 1957. "Groups and conditional Monte Carlo". Ann. Math. Stat.28: 1048-1052.

[40] H.O.Madsen, S. Krenk, N.C. Lind, "Methods of Structural Safety", 1986, Prentice Hall.

[41] Maes, M.A., "Tail Heaviness in Structural Reliability", Proc. CERRA-ICASP, pp997-1002, 1995, Paris.

[42] Maes, M.A. and Breitung, K., 1993,"Reliability-Based Tail Estimation", Proc. IUTAM Symposium on Probabilistic Mechanics, pp335-346, San Antonio, Texas, USA, 1993

[43] Breiman, L., Stone, C.J., and Gins, J.D., 1979, "New Methods for Estimating Tail Probabilities and Extreme Value Distributions", TSCorp, PD-A226-1, Santa Monica, California.

[44] Hosking, J.R.M. and Wallis, J.R., 1987, "Parameter and quantile estimation for the generalised Pareto distribution", Technometrics, 29(3): 339-349.

[45] Castillo, E., 1988, "Extreme Value Theory in Engineering", Academic Press, San Diego, California.

[46] Greenwood, J.A., Landwehr, J.M., Matalas, N.C. and Wallis, J.R., 1979. "Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form". Water Resour. Res., 15(5): 1049-1054.

[47] Pickands, J., "Statistical Interference Using Extreme Order Statistics", Annals of Statistics, Vol. 3, pp.119-113, 1975

[48] Kendall, M., and Stuart, A., "The Advanced Theory of Statistics", Vol. II, Fourth Edition, London, 1979

[49] Ashkar, F., Ouarda, T.B.M.J, "On some methods of fitting the Generalised Pareto distribution", J. Hydrology 177(1996), pp117-141

[50] Ashkar, F., Ouarda, T.B.M.J. and Hache, M., 1994. "The generalised method of moments for filling the generalised Pareto distribution", Scientific Report STAT-14, Department of Mathematics, University of Moncton, N.H..

[51] Yolanda Carson, Anu Maria, "Simulation Optimisation: Methods and Applications", 1997 winter Simulation Conference.

[52] Hosking, J.R.M., 1990. "L-Moments: analysis and estimation of distributions using linear combination of order statistics". R. Stat. Soc., Ser. B, 52(l): 105-124.

[53] Boos, D.D., 1984, "Using Extreme Value Theory to Estimate Large Percentiles", Technometrics, Vol 2, No. 1, pp. 33-39.

[54] Kendall, M.G. and Stuart, A., 1963. "The Advanced Theory of Statistics", Vol. 1. Charles Griffin, London.

[55] Smith, J.A., 1987. "Estimating the upper tail of flood frequency distributions". Water Resour. Res., 23(8): 1657-1666.

[56] Van Montfort, M.A.J. and Witter, J.V., 1986. "The generalised Pareto distribution applied to rainfall depths", Hydrol. Sci. J., 31(2): 151-162.

[57] Tamano, T. et al, "A New Empirical Formula for Collapse Resistance of Commercial Casing", ASME Transactions of Energy Resources Technology, pp.489-495, 1983

[58] Issa, J. A et al, "An Improved Design Equation For Tubular Collapse", SPE 26317,Proceedings of SPE Annual Technical Conference, Houseton, OCT.,1993

[59] Tokimasa, K. et al, "FEM Analysis of The Collapse Strength of A Tube", J. of Pressure Vessel Technology, ASME, May, 1986.

[60] Mantz, l.S. and Munro, A.H., "On the Use of the Generalised Extreme Value Distribution in Estimating Extreme Quantiles", Biometrics, 23, (March., 1967),79-103.

[61] Weissman, I., "Estimation of Parameters and Large Quantiles Based on the k Largest Observations", J. Am. Statistical Assn., 73,264, (1978),812-815.

[62] Hasofer, A.M. and Wang, Z., "System Reliability Calculations Using Extreme Value Theory", Proc. CASP, I, (1991),41-47.

[63] Kendall, M. and Stuart.A., "The Advanced Theory of Statistics", Volume II, Fourth Edition, MacMillan, London, 1979.7.

[64] Berger, l.O., "Statistical Decision Theory and Bayesian Analysis", Second Edition, Springer, New York, 1985.

[65] Box, G.E.P. and Muller, M. E.: "A Note on the Generation of Random Normal Deviates". Annals of Math Statistics, Vol. 29, 1958.

[66] Kotz, S. and Johnson, N.L. (Editors), 1985. "Pareto distribution". Encyclopaedia of Statistical Sciences, Vol. 6. Wiley, New York, pp. 568-574.

[67] Bulletin on formulas and calculations for casing, tubing, drill pipe and line pipe properties, API Bulletin 5C3, 6th edition, Oct 1994.

[68] DNV Software Report: SESAM Theory Manual: PROBAN, Det Norske Veritas,1996

[69] SESAM Theory Manual of PROBAN, "General Purpose probability analysis program", 1996

[70] McGuinness, P., 1995, "Risk Assessment: A line Manager's Guide", Bourne Press

[71] Vose, Davide, "Quantitative Risk Analysis: A Guide to Monte Carlo Simulation Modelling", John Willey & Sons, 1996

[72] Adams, "An investigation into the Application of QRA in Casing Design", SPE48319, Society of Petroleum Engineers, 1998

[73] A.G. Tallin, P.R. Paslay, "The Development of Risk-Based Burst Design for Well Casing and Tubing", SPE48320, Society of Petroleum Engineers, 1998

[74] M.L. Payne, U.B. Sathuvalli, "Select Topics and Applications of Probabilistic OCTG Design", SPE48324, Society of Petroleum Engineers, 1998

[75] Richard A. Miller, "Real World Implementation of QRA Methods in Casing Design", SPE48325, Society of Petroleum Engineers, 1998

[76] Andy Hinton, "Will Risk Based Casing Design Mean Safer Wells?", SPE48326, Society of Petroleum Engineers, 1998

[77] C.V. Burres, A.G. Tallin, "Determination of Casing and Tubing Burst and Collapse Design Factors to Achive Target Levels of Risk, Including Influence of Mill Source", SPE48321, Society of Petroleum Engineers, 1998

[78] James B. Raney, P.V.R. Suryanarayana, "A Comparison of Deterministic and Reliability-Based Design Methodologies for Production Tubing", SPE48322, Society of Petroleum Engineers, 1998

[79] A.J. Adams, A.V.R. Warren, " On the Development of Reliability-Based Design Rules for Casing Collapse", SPE48331, Society of Petroleum Engineers, 1998

[80] M.H. Aldin, K. Logan, "A System for Comprehensive Measurements of Pipe Wall Thickness and Diameter in Support of Risk-Based Tubular Design", SPE48333, Society of Petroleum Engineers, 1998

[81] A.P. Assanelli, R.G. Toscano and D.H. Johnson, "Collapse Behaviour of Casings: Measurement Techniques, Numerical Analyses and Full Scale Testing", SPE/ATW, Society of Petroleum Engineers, 1998

[82] G.T. Ju, T.L. Power and A.G. Tallin, " Reliability Approach to the Design of OCTG Tubulars Against Collapse", SPE48332, Society of Petroleum Engineers, 1998

[83] J.L. Thorogood and T.W. Hogg, "Application of Risk Analysis Methods to Subsurface well Collisions", SPE Drilling Engineering, 1991 6(4) PP299-304

[84] YoLanda Carson and Anu Maria, "Simulation Optimisation: Methods and Applications", Proceedings of the 1997 Winter Simulation Conference, pp118-126

[85] Eduardo Saliby, "Descriptive Sampling: An Improvement Over Latin Hypercube Sampling", Proceedings of the Winter Simulation Conference, pp230-233

[86] S. Nadarajah, "Simulation of Multivariate Extreme Values", Proceedings of the 1997 Winter Simulation Conference, pp281-285

[87] J.O. Miller, "How Common Random Numbers Affect Multi-normal Selection", Proceedings of the 1997 Winter Simulation Conference, pp342-347

[88] Hisham A. Al-Mharmah and James M. Calvin, "Comparison of Monte Carlo and Deterministic Methods for Non-adaptive Optimisation", Proceedings of the 1997 Winter Simulation Conference, pp348-351

[89] P.Bjerager, "On Computation Methods for Structure Reliability Analysis" Structural Safety, 9 (1990), pp77-96

[90] R.E. MelChers, "Structural Reliability Analysis and Prediction", Wiley, New York, 1987

[91] R.Y. Rubinstein, "Simulation and the Monte Carlo Method", Wiley, NewYork, 1981

[92] M. Shinozuka, "Basic Analysis of Structural Safety", J. Struct. Eng., 109(1983), pp721-740

[93] A. Harbitz, "Efficient and Accurate Probability of Failure Calculation by the Use of Importance Sampling Technique", Proc. 4th International Conference on Application of Statistics and Probability in Soil and Structural Engineering, ICASP4, Pitagora, Italy, 1983, pp825-886

[94] V. Bourgund and C.G. Buther, "Importance Sampling Procedure Using Design Point", ISPUD, Institute of Engineering and Mechanics, Innsbruck, Austria, Report 9-86, 1996

[95] A. Karamchandani, P. Bjerager and R.A. Cornell, " Adaptive Importance Sampling", Proc. 5th Int. Conf. On Structural Safety and Reliability, 1989, Vol. II, pp855-862

[96] R.E. Melchers, "Improved Importance Sampling Methods for Structural System Reliability Calculation", Proc. 5th int. Conf. On Structural Safety and Reliability, San Francisco, Vol.2, 1989, pp1185-1192

[97] G.L. Ang, A.H.-S. Ang and W.H. Tang, "Kernel Method in Importance Sampling Density Estimation", Proc. 5th int. Conf. On Structural Safety and Reliability, San Francisco, Vol.2, 1989, pp1193-1200

[98] M. Hohenbichler and R. Rackwitz, "Improvement of Second-Order Reliability Estimates by Importance Sampling", J. Eng. Mechs., 114(12) (1998), pp2159-2199

[99] K. Breitung, "Asymptotic Approximations for Multi-normal Integrals", J. Eng. Mechs., 110(3) (1984), pp357-366

[100] J.Galambos, "The Asymptotic Theory of Extreme Order Statistics", Wiley, New York, 1978

[101] P.geyskens, A. Der Kiureghian and G. De Roeck, "SORM Analysis Using Quasi-Newton Optimisation", Proc. Int. Conf. On Computational Stochastic Mechs., Corfu, Greece, September 1991

[102] R.E. Melchers, "Simulation in Time-invariant and Time-variant Reliability Problems", Proc. 4th IFIP WG7.5 Conf.: Reliability and Optimisation of Systems, Munich, 1991, pp37-82

[103] Jerry Banks, "The Future of Simulation Software: A Panel Discussion", Proceedings of the 1997 Winter Simulation Conference, pp166-173

[104] Hussain Rabia, "Fundamentals of Casing Design", Graham & Trotman, 1987

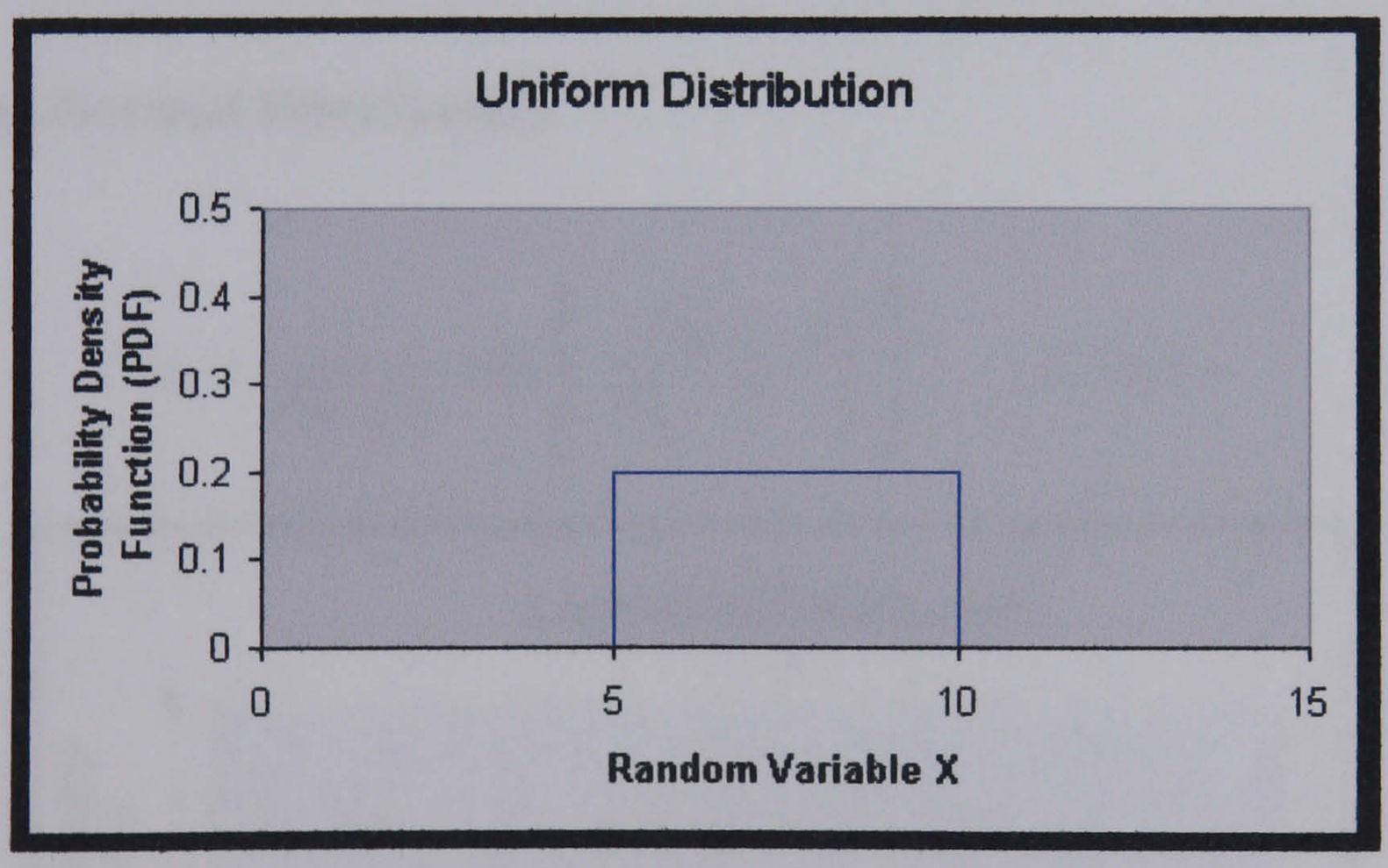
[105] Husain Rabia, "Oilwell Drilling Engineering", Graham & Trotman, 1986

Chapter 11 : Appendix

Appendix A: Most Frequently Used Distribution Types

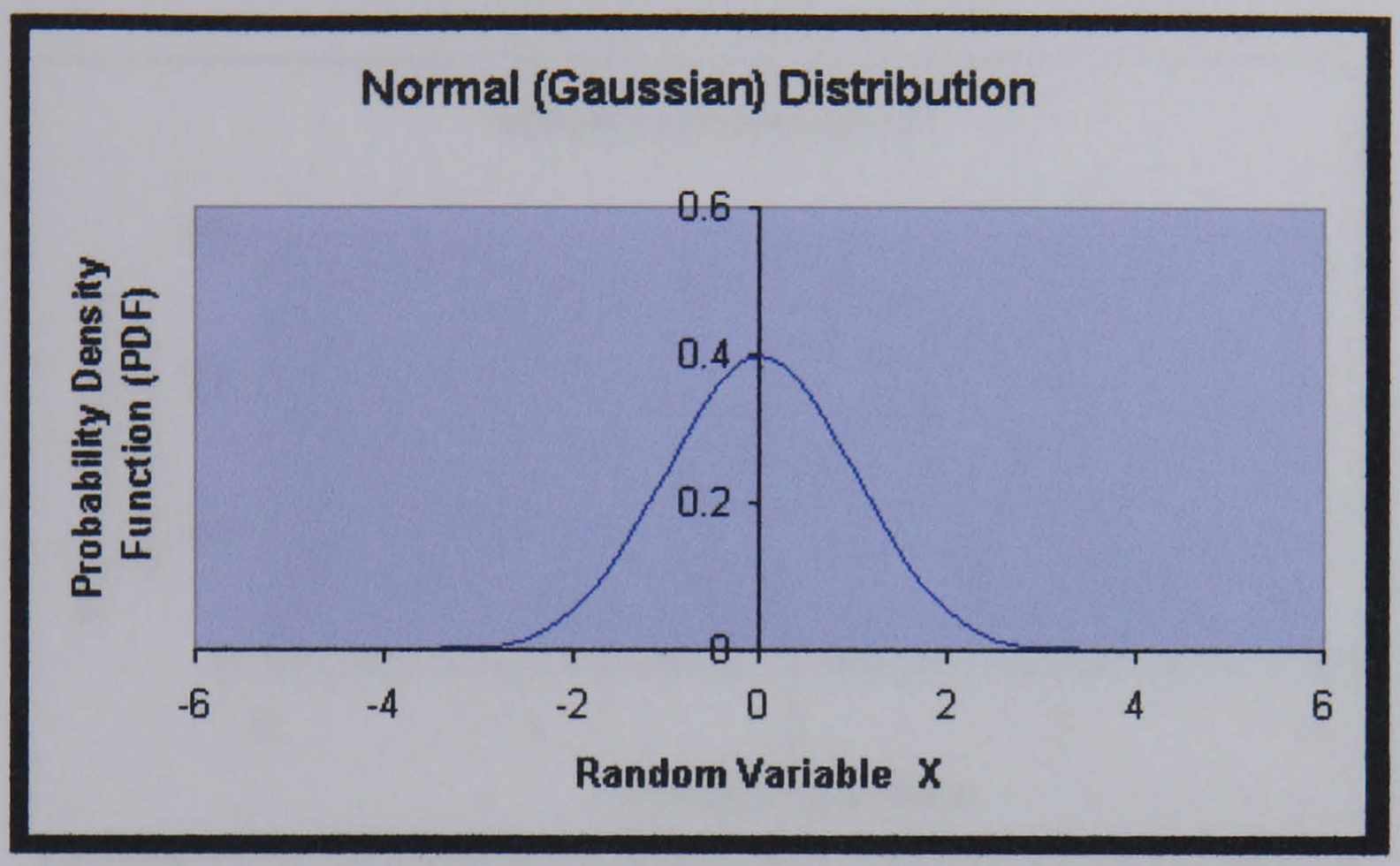
A.1 Uniform Distribution

In a uniform distribution, all the values within a range are equally likely. We randomly choose a number between 0 and 1 (uniformly distributed) and transform it to the desired range.



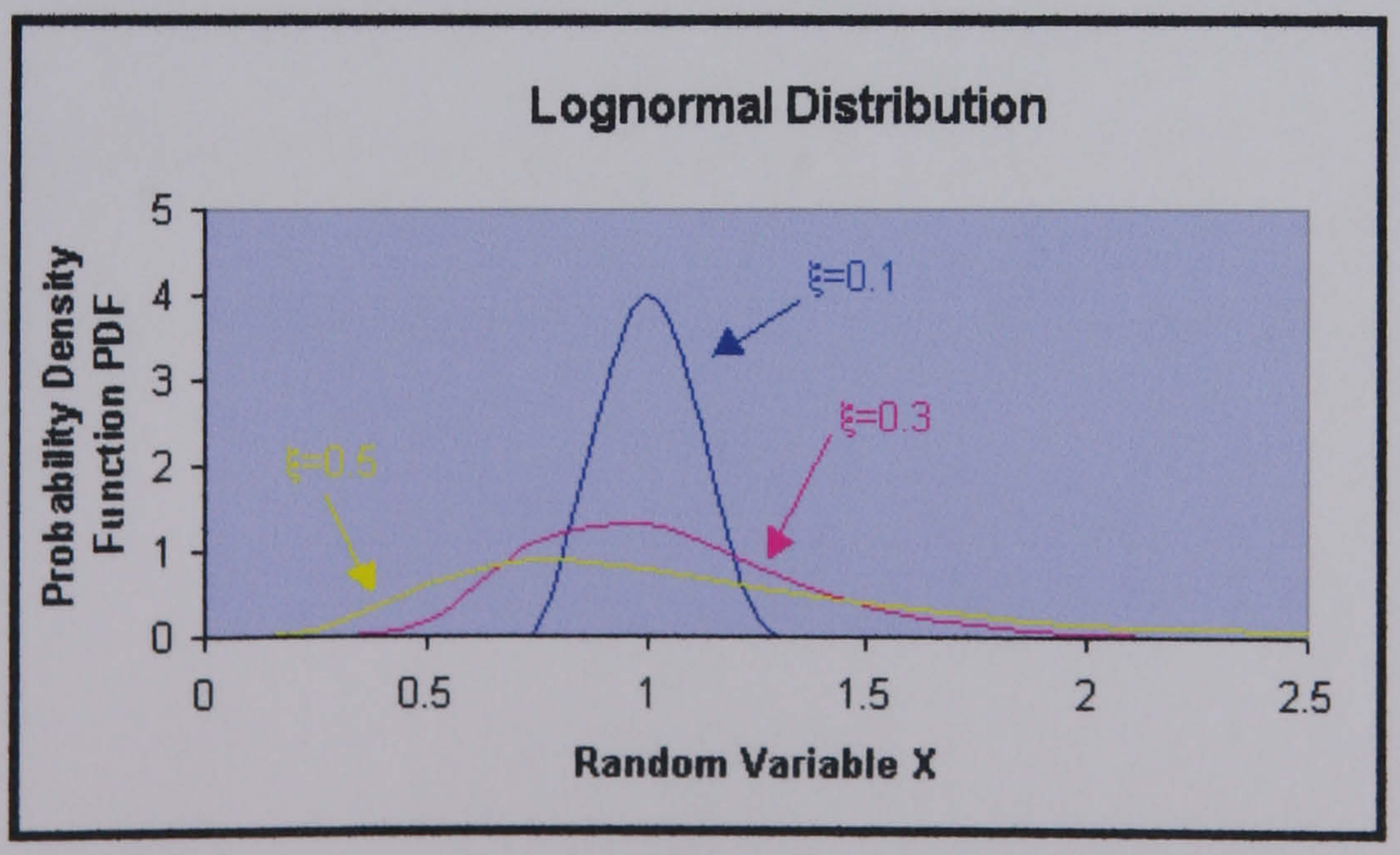
A.2 Normal Distribution

$$f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \qquad -\infty \le x \le \infty$$



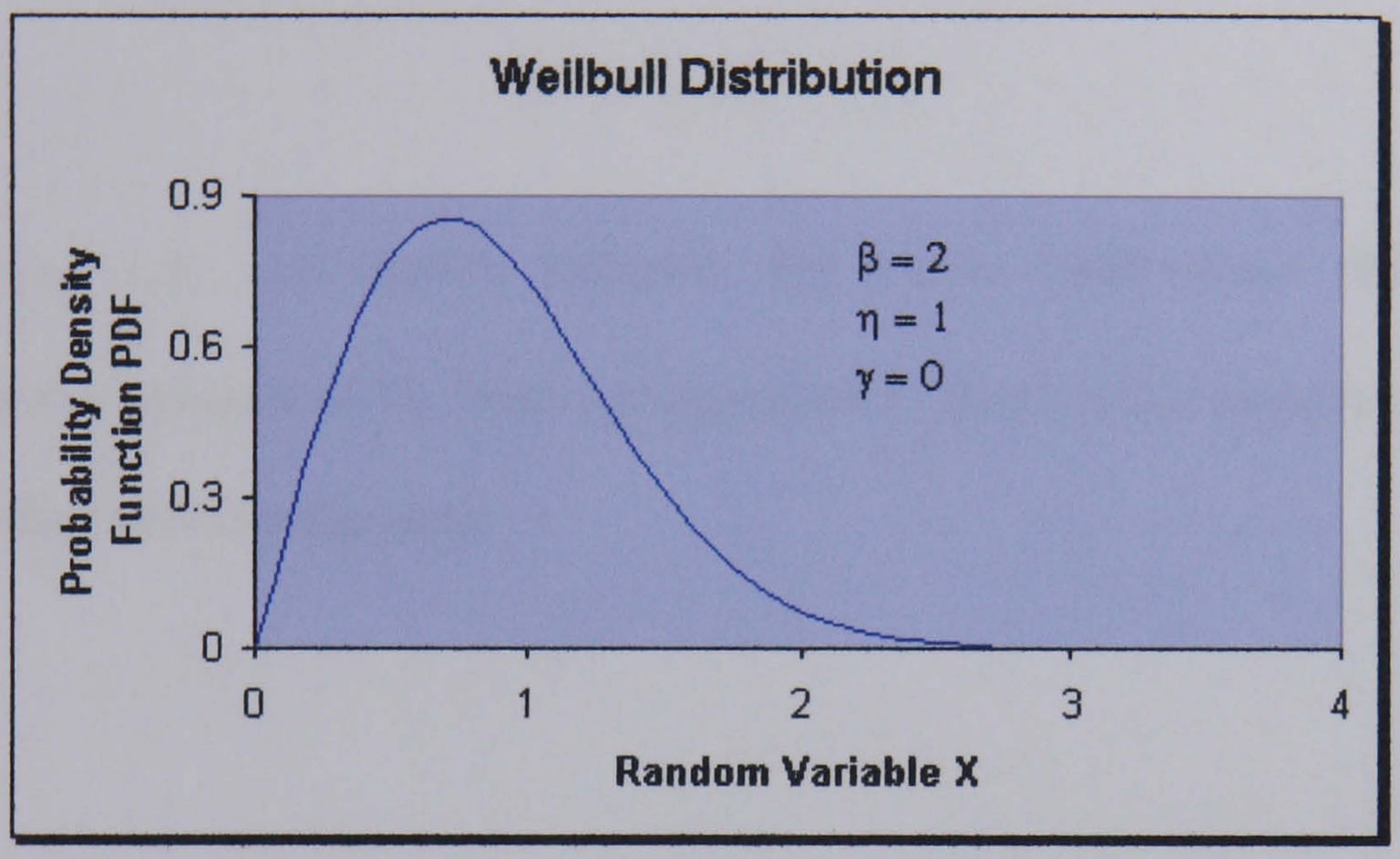
A.3 Log-Normal Distribution

$$f_x = \frac{1}{\sqrt{2\pi}\,\zeta\,x} \exp\left[-\frac{1}{2}\left(\frac{\ln x - \lambda}{\zeta}\right)^2\right] \qquad 0 \le x < \infty$$



A.4 Weibull Distribution

$$f(x) = \frac{\beta}{\eta}\left(\frac{x-\gamma}{\eta}\right)^{\beta-1} \exp\left[-\left(\frac{x-\gamma}{\eta}\right)^{\beta}\right]$$



Appendix B: Algorithm of Monte Carlo Simulation

Suppose the probability model:

$$M = g(x_1, x_2, \cdots, x_n)$$

where $x_1, x_2, \cdots, x_n$ are random variables with known distributions. We want to determine the statistics of M. Then the algorithm of Monte Carlo Simulation can be accomplished as following steps.

Steps:

1. In each run, randomly generate the values of $x_1, x_2, \cdots, x_n$ according to their statistics.

2. Calculate $M = g(x_1, x_2, \cdots, x_n)$ for each run.

3. Repeat steps 1 and 2 N times.

4. Compute the statistics of M with the N samples obtained in steps 1 to 3.

Appendix C: Random Number Generators

C.1 General

Because simulation studies and Monte Carlo analysis has to be made of long sequences of random numbers (generally pseudo-random numbers). These are most conveniently generated using a digital computer. The increased use of such studies in recent years has meant the reliable library functions have been made available on most computer systems. It can normally be assumed that these functions have been fully tested for random behaviour.

C.2 Uniform Random Number Generators

Most digital random number generators are based on uniform pseudo-random number generators of the multiple congruence type. A uniform random number generator is one which generates successive independent realisations u_i of a random variable U having a rectangular density function, usually in the interval [0,1], i.e.

$$f_U(u) = \begin{cases} 1 & 0 \le u \le 1 \\ 0 & \text{elsewhere} \end{cases} \tag{C-1}$$

giving

$$f_U(u) = \begin{cases} 1 & u < 0 \\ u & 0 \le u \le 1 \\ 1 & u > 1 \end{cases} \tag{C-2}$$

the mean and standard deviation of the random variable U can be shown to be

$$\mu_U = 0.5 \qquad \sigma_U = \sqrt{1/12} \tag{C-3}$$

C.3 Multiple Congruence Method

This method produces a series of pseudo-random numbers r_i that eventually repeats. But, if correctly designed, only after a very long cycle, the next number in the pseudo-random series is related to the previous number by the relationship

$$r_{n+1} = a\,r_n (\text{modulo } m) \tag{C-4}$$

where a and m are integer constants and are relatively prime. Starting with an integer seed i_0, the first pseudo-random number r_1 in the interval [0,1] is obtained from

$$\frac{ai_0}{m} = jl1 + \frac{ai_0 - j_1 m}{m} = j_1 + \frac{i_1}{m} = j_1 + r_1 \tag{C-5}$$

i.e. r_1 is the fractional part of the quotient $(ai_0)/m$ and j_1 is the integer part. Then, $i_1 = (ai_0 - j_1 m)$ is the seed for the second random number.

In general, the integer constants a and m are chosen to obtain the longest possible cycle. It can be shown that if

$$m = 2^b \quad \text{and} \quad a = (8t \pm 3) \tag{C-6}$$

where b and t are integers. Then the length of the integer sequence before repetition is of the order of $2^{(b-2)}$.

The pseudo-random numbers r_i generated by this method can be assumed to be independent realisations u_i of a random variable U having a rectangular distribution with $0 \leq U \leq 1$.

C.4 Generation of Random Deviations with a Specified Probability Distribution Function F_X

A convenient general method consist of generating a random number r, as described above and then, by making use of equation (C-2), finding the corresponding random deviate x of the random variable X from

$$F_X(x) = F_U(u) = u = r \tag{C-7}$$

where F_X is the required distribution function. It is therefore necessary to find the inverse function F^{-1}, giving

$$x = F_X^{-1}(r) \tag{C-8}$$

This is valid for all forms of distribution function, but two classes of variable exist which require different treatment.

Class A: The distribution function F_X has an inverse F_X^{-1} which can be expressed in closed form.

In this case the random deviate x can be generated simply by obtain successive values

$$x_i = F_X^{-1}(r_i) \tag{C-9}$$

Class B: The distribution function F_X has an inverse F_X^{-1} which cannot be expressed in closed form

In this case the general procedure is the same, but the inverse function has to be evaluated either graphically, by numerical integration by table look-up and interpolation, or by fitting an appropriate polynomial.

C.5 Special Cases: Generation of Random Deviations Having Normal and Log-normal Distributions

The normal and log-normal distributions are two of the Class B functions, but because of their frequent use they deserve further attention. In addition to the general method described above, a number of special methods exist for normal variables. These methods may also be used for generating log-normally distributed random numbers, by the use of an appropriate transformation.

- *Generation of random normal deviates from the sum of n rectangular distributed random deviates.*

The fact that under very general conditions, the distribution functions for the sum of series of independent random variables tends to a normal distribution as the number of variables in the sum of increases can be used to generate random numbers having a distribution which approximates very closely to normal.

Most computer routines use the sum of 12 or more independent rectangular distributed random numbers r_i. If the latter are generated in the interval [1, a], their sum ξ can easily be shown to be approximately normally distributed with mean μ_ξ given by

$$\mu_\xi = an/2 \tag{C-10}$$

and variance σ_ξ^2 by

$$\sigma_\xi^2 = a^2 n/12 \tag{C-11}$$

for the simple case when a=1 and n=12, ξ' given by

$$\xi' = \sum_{i=1}^{12} r_i - 6 \tag{C-12}$$

is approximately normally distributed with zero mean and unit standard deviation.

This approach given excellent approximations to the normal distribution for deviates within two or three standard deviations from the mean, but for extreme values the approximation becomes increasingly poor, unless n is large. For example, the random variable ξ' defined by Eq(C-12) cannot lie outside the interval [-6, 6].

- *Generation of random normal deviates using method due to Box and Muller*

Box and Buller[65] have shown that if r_1 and r_2 are in dependent random variables from the same rectangular distribution in the interval [0, 1], then N_1 and N_2 given by

$$N_1 = (-2\ln r_1)^{\frac{1}{2}} \cos 2\pi r_2 \tag{C-13a}$$

$$N_2 = (-2\ln r_1)^{\frac{1}{2}} \sin 2\pi r_2 \tag{C-13b}$$

are independent random variables, normally distributed with zero mean and unit variance.

The advantage of this method is that it is accurate over the complete range and depends only on the randomness and independence of r_1 and r_2.

Appendix D: Casing Properties

Table TD-1: Properties of $13\frac{1}{2}$" casings

1	2	3	4	5	6	7	8	9	10	11	12
Size: outside diameter D in mm	Nominal weight, threads and coupling lb/ft kg/m	Grade	Pipe		Threaded and coupled			Extreme-line		Collapse resistance P_c psi bar	Pipe body yield strength P_y 1000 lb kN
			Wall thickness t	Inside diameter d	Drift diameter	Regular	Special clearance	Drift diameter	Outside diameter W		
						Outside diameter					
						W	W_c				
						in mm					
13⅜ 339.7	54.50 81.2	J-55	0.380 9.65	12.615 320.4	12.459 316.46	14.375 365.1				1130 78	853 3794
	61.00 90.9	J-55	0.430 10.92	12.515 317.9	12.359 313.92	14.375 365.1				1540 106	962 4279
	68.00 101	J-55	0.480 12.19	12.415 315.3	12.259 311.36	14.375 365.1				1950 134	1069 4755
	72.00* 107	J-55	0.514 13.06	12.347 313.6	12.191 309.65	14.375 365.1				2230 154	1142 5080
	54.50 81.2	K-55	0.380 9.65	12.615 320.4	12.459 316.46	14.375 365.1				1130 78	853 3794
	61.00 90.9	K-55	0.430 10.92	12.515 317.9	12.359 313.92	14.375 365.1				1540 106	962 4279
	68.00 101	K-55	0.480 12.19	12.415 315.3	12.259 311.36	14.375 365.1				1950 134	1069 4755
	72.00* 107	K-55	0.514 13.06	12.347 313.6	12.191 309.65	14.375 365.1				2230 154	1142 5080
	61.00* 90.9	C-75	0.430 10.92	12.515 317.9	12.359 313.92	14.375 365.1				1660 114	1312 5836
	68.00 101	C-75	0.480 12.19	12.415 315.3	12.259 311.36	14.375 365.1				2220 153	1458 6486
	72.00 107	C-75	0.514 13.06	12.347 313.6	12.191 309.65	14.375 365.1				2590 179	1656 6930
	61.00* 90.9	L-80	0.430 10.92	12.515 317.9	12.359 313.92	14.375 365.1				1670 115	1399 6223
	68.00 101	L-80	0.480 12.19	12.415 315.3	12.259 311.36	14.375 365.1				2260 156	1656 6921
	72.00 107	L-80	0.514 13.06	12.347 313.6	12.191 309.65	14.375 365.1				2670 184	1661 7388
	61.00* 90.9	N-80	0.430 10.92	12.515 317.9	12.359 313.92	14.375 365.1				1670 115	1399 6223
	68.00 101	N-80	0.480 12.19	12.415 315.3	12.259 311.36	14.375 365.1				2260 156	1656 6921
	72.00 107	N-80	0.514 13.06	12.347 313.6	12.191 309.65	14.375 365.1				2670 184	1661 7388
	61.00* 90.9	MW-C-80	0.430 10.92	12.515 317.9	12.359 313.92	14.375 365.1				1670 115	1574 7002
	68.00* 101	MW-C-80	0.480 12.19	12.415 315.3	12.259 311.36	14.375 365.1				2320 160	1750 7784
	72.00* 107	MW-C-80	0.514 13.06	12.347 313.6	12.191 309.65	14.375 365.1				2760 192	1868 8314
	61.00* 90.9	C-95	0.430 10.92	12.515 317.9	12.359 313.92	14.375 365.1				1670 115	1661 7388
	68.00 101	C-95	0.480 12.19	12.415 315.3	12.259 311.36	14.375 365.1				2320 160	1847 8216
	72.00 107	C-95	0.514 13.06	12.347 313.6	12.191 309.65	14.375 365.1				2820 194	1973 8776
	61.00* 90.9	MW-C-95	0.430 10.92	12.515 317.9	12.359 313.92	14.375 365.1				1670 115	1661 7388
	68.00* 101	MW-C-95	0.480 12.19	12.415 315.3	12.259 311.36	14.375 365.1				2320 160	1847 8216
	72.00* 107	MW-C-95	0.514 13.06	12.347 313.6	12.191 309.65	14.375 365.1				3470 239	1973 8776
	61.00* 90.9	P-110	0.430 10.92	12.515 317.9	12.359 313.92	14.375 365.1				1670 115	1924 8558
	68.00* 101	P-110	0.480 12.19	12.415 315.3	12.259 311.36	14.375 365.1				2330 161	2136 9515
	72.00* 107	P-110	0.514 13.06	12.347 313.6	12.191 309.65	14.375 365.1				2890 199	2284 10 160
	68.00* 101	MW-125	0.480 12.19	12.415 315.3	12.259 311.36	14.375 365.1				2330 161	2481 10 814
	72.00* 107	MW-125	0.514 13.06	12.347 313.6	12.191 309.65	14.375 365.1				2890 199	2696 11 548
	68.00* 101	MW-140	0.480 12.19	12.415 315.3	12.259 311.36	14.375 365.1				2330 161	2722 12 106
	72.00* 107	MW-140	0.514 13.06	12.347 313.6	12.191 309.65	14.375 365.1				2890 199	2907 12 931
	68.00* 101	V-150	0.480 12.19	12.415 315.3	12.259 311.36	14.375 365.1				2330 161	2917 12 975
	72.00* 107	V-150	0.514 13.06	12.347 313.6	12.191 309.65	14.375 365.1				2890 199	3115 13 856
	68.00* 101	MW-155	0.480 12.19	12.415 315.3	12.259 311.36	14.375 365.1				2330 161	3014 13 407
	72.00* 107	MW-155	0.514 13.06	12.347 313.6	12.191 309.65	14.375 365.1				2890 199	3219 14 319

* Non-API.

Table TD-1 *Continued*

13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Internal yield pressure at minimum yield							Joint strength							
Plain-end or extreme-line	Round thread		Buttress/ BDS/ Omega				Threaded and coupled						Extreme-line	
	Short	Long[a]	Regular coupling		Special clearance coupling		Round thread		Buttress/ BDS/ Omega					
							Short	Long	Regular coupling		Special clearance coupling			
			Same grade	Higher grade	Same grade	Higher grade			Same grade	Higher grade	Same grade	Higher grade	Standard joint	Optional joint
P_i psi							P_j 1000 lb							
bar							kN							
2730	2730	2730	2730	2730			614	596	908	909				
188	188	188	188	188			2286	2651	4043	4043				
3000	3000	3000	3000	3000			596	690	1025	1025				
213	213	213	213	213			2647	3069	4559	4559				
3450	3450	3450	3450	3450			675	783	1140	1140				
238	238	238	238	238			3003	3483	5071	5071				
3700	3700		3700	3700			720		1218	1218				
255	255		255	255			3247		5418	5418				
2730	2730	2730	2730	2730			547	636	1038	1038				
188	188	188	188	188			2433	2829	4617	4617				
3000	3000	3000	3000	3000			633	736	1166	1166				
213	213	213	213	213			2816	3274	5200	5200				
3450	3450	3450	3450	3450			718	836	1300	1300				
238	238	238	238	238			3194	3714	5783	5783				
3700	3700		3700	3700			776		1389	1389				
255	255		255	255			3452		6179	6179				
4220	4220		4220	4220			786		1345	1345				
291	291		291	291			3550		5983	5983				
4710	4710		4710	4710			906		1466	1466				
325	325		325	325			4030		6655	6655				
5040	5040	5040	5040	5040			978	1133	1586	1586				
347	347	347	347	347			4350	5040	7108	7108				
4500	4500		4500	4500			836		1366	1366				
310	310		310	310			3732		6179	6179				
5020	5020		5020	5020			952		1545	1545				
346	346		346	346			4235		6873	6873				
5390	5390	5390	5390	5390			1029	1181	1650	1650				
371	371	371	371	371			4577	5298	7340	7340				
4500	4500		4500	4500			849		1426	1426				
310	310		310	310			3777		6343	6343				
5020	5020		5020	5020			963		1585	1585				
346	346		346	346			4284		7050	7050				
5390	5390	5390	5390	5390			1040	1205	1663	1663				
371	371	371	371	371			4626	5360	7531	7531				
5060	5060		5060	5060			931		1514	1514				
349	349		349	349			4141		6735	6735				
5650	5650		5650	5650			1057		1663	1663				
390	390		390	390			4702		7486	7486				
6050	6050		6050	6050			1142		1796	1796				
417	417		417	417			5080		7986	7986				
5340	5340		5340	5340			962		1564	1564				
368	368		368	368			4368		7090	7090				
5870	5870		5870	5870			1114		1772	1772				
412	412		412	412			4955		7882	7882				
6390	6390	6390	6390	6390			1204	1362	1863	1863				
441	441	441	441	441			5356	6192	8420	8420				
5340	5340		5340	5340			991		1630	1630				
368	368		368	368			4408		7251	7251				
5870	5870		5870	5870			1125		1812	1812				
412	412		412	412			5004		8060	8060				
6390	6390		6390	6390			1215		1935	1935				
441	441		441	441			5405		8607	8607				
6190	6190		6190	6190			1143		1670	1670				
427	427		427	427			5084		8318	8318				
6910	6910		6910	6910			1297		2079	2079				
476	476		476	476			5769		9248	9248				
7400	7400		7400	7400			1402		2221	2221				
510	510		510	510			6236		9980	9980				
7850	7850		7850	7850			1466		2306	2306				
541	541		541	541			6490		10 258	10 258				
8410	8410		8410	8410			1577		2463	2463				
580	580		580	580			7015		10 956	10 956				
8780	8780		8780	8780			1632		2573	2573				
606	606		606	606			7269		11 445	11 445				
9420	9420		9420	9420			1763		2748	2748				
649	649		649	649			7842		12 228	12 228				
9420	9420		9420	9420			1747		2752					
649	649		649	649			7771		12 242					
10 000	10 000		10 000	10 000			1887		2969					
690	690		690	690			8394		13 073					
8730	8730		8730	8730			1894		2841					
671	671		671	671			8025		12 637					
10 420	10 420		10 420	10 420			1948		2894					
718	718		718	718			8670		13 496					

Table TD-2: Properties of $9\frac{5}{8}$" casings

1	2	3	4	5	6	7	8	9	10	11	12
Size: outside diameter D *in* *mm*	*Nominal weight, threads and coupling* *lb/ft* *kg/m*	*Grade*	*Pipe*		*Threaded and coupled*			*Extreme-line*		*Collapse resistance* P_c *psi* *bar*	*Pipe body yield strength* P_y *1000 lb* *kN*
			Wall thickness t	*Inside diameter d*	*Drift diameter*	*Regular*	*Special clearance*	*Drift diameter*	*Outside diameter W***		
						Outside diameter					
						W	W_c				
						in *mm*					
$9\frac{5}{8}$	36.00	J-55	0.352	8.921	8.765	10.625	10.125			2020	564
244.5	53.6		8.94	226.6	222.63	269.9	257.2			139	2509
	40.00	J-55	0.395	8.835	8.679	10.625	10.125	8.599	10.100	2570	630
	59.6		10.03	224.4	220.45	269.9	257.2	218.41	256.5	177	2802
	36.00	K-55	0.362	8.921	8.765	10.625	10.125			2020	564
	53.6		8.94	226.6	222.63	269.9	257.2			139	2509
	40.00	K-55	0.395	8.835	8.679	10.625	10.125	8.599	10.100	2570	630
	59.6		10.03	224.4	220.45	269.9	257.2	218.41	256.5	177	2802
	40.00	C-75	0.395	8.835	8.679	10.625	10.125	8.599	10.100	2990	859
	59.6		10.03	224.4	220.45	269.9	257.2	218.41	256.5	205	3821
	43.50	C-75	0.435	8.755	8.599	10.625	10.125	8.599	10.100	3750	942
	64.8		11.05	222.4	218.41	269.9	257.2	218.41	256.5	259	4190
	47.00	C-75	0.472	8.681	8.525	10.625	10.125	8.525	10.100	4630	1018
	70.0		11.99	220.5	216.54	269.9	257.2	216.54	256.5	319	4528
	53.50	C-75	0.545	8.535	8.379	10.625	10.125	8.379	10.100	5360	1166
	79.7		13.84	216.8	212.83	269.9	257.2	212.83	256.5	440	5187
	58.40*	C-75	0.595	8.435	8.279	10.625	10.125			7570	1286
	87.0		15.11	214.2	210.30	269.9	257.2			522	5631
	40.00	L-80	0.395	8.835	8.679	10.625	10.125	8.599	10.100	3090	916
	59.6		10.03	224.4	220.45	269.9	257.2	218.41	256.5	213	4075
	43.50	L-80	0.435	8.755	8.599	10.625	10.125	8.599	10.100	3810	1005
	64.8		11.05	222.4	218.41	269.9	257.2	218.41	256.5	263	4470
	47.00	L-80	0.472	8.681	8.525	10.625	10.125	8.525	10.100	4750	1086
	70.0		11.99	220.5	216.54	269.9	257.2	216.54	256.5	328	4831
	53.50	L-80	0.545	8.535	8.379	10.625	10.125	8.379	10.100	6620	1244
	79.7		13.84	216.8	212.83	269.9	257.2	212.83	256.5	456	5534
	58.40*	L-80	0.595	8.435	8.279	10.625	10.125			7860	1350
	87.0		15.11	214.2	210.30	269.9	257.2			544	6005
	40.00	N-80	0.395	8.835	8.679	10.625	10.125	8.599	10.100	3090	916
	59.6		10.03	224.4	220.45	269.9	257.2	218.41	256.5	213	4075
	43.50	N-80	0.435	8.755	8.599	10.625	10.125	8.599	10.100	3810	1005
	64.8		11.05	222.4	218.41	269.9	257.2	218.41	256.5	263	4470
	47.00	N-80	0.472	8.681	8.525	10.625	10.125	8.525	10.100	4750	1086
	70.0		11.99	220.5	216.54	269.9	257.2	216.54	256.5	328	4831
	53.50	N-80	0.545	8.535	8.379	10.625	10.125	8.379	10.100	6620	1244
	79.7		13.84	216.8	212.83	269.9	257.2	212.83	256.5	456	5534
	58.40*	N-80	0.595	8.435	8.279	10.625	10.125			7890	1350
	87.0		15.11	214.2	210.30	269.9	257.2			544	6005
	40.00*	MW-C-90	0.395	8.835	8.679	10.625	10.125	8.599	10.100	3260	1031
	59.6		10.03	224.4	220.45	269.9	257.2	218.41	256.5	225	4586
	43.50*	MW-C-90	0.435	8.755	8.599	10.625	10.125	8.599	10.100	4010	1130
	64.8		11.05	222.4	218.41	269.9	257.2	218.41	256.5	276	5026
	47.00*	MW-C-90	0.472	8.681	8.525	10.625	10.125	8.525	10.100	4990	1222
	70.0		11.99	220.5	216.54	269.9	257.2	216.54	256.5	344	5436
	53.50*	MW-C-90	0.545	8.535	8.379	10.625	10.125	8.379	10.100	7110	1399
	79.7		13.84	216.8	212.83	269.9	257.2	212.83	256.5	490	6223
	58.40*	MW-C-90	0.595	8.435	8.279	10.625	10.125			8580	1519
	87.0		15.11	214.2	210.30	269.9	257.2			590	6757
	40.00*	C-95	0.395	8.835	8.679	10.625	10.125	8.599	10.100	3330	1088
	59.6		10.03	224.4	220.45	269.9	257.2	218.41	256.5	230	4840
	43.50*	C-95	0.435	8.755	8.599	10.625	10.125	8.599	10.100	4130	1193
	64.8		11.05	222.4	218.41	269.9	257.2	218.41	256.5	285	5307
	47.00*	C-95	0.472	8.681	8.525	10.625	10.125	8.525	10.100	5090	1289
	70.0		11.99	220.5	216.54	269.9	257.2	216.54	256.5	350	5734
	53.50*	C-95	0.545	8.535	8.379	10.625	10.125	8.379	10.100	7330	1477
	79.7		13.84	216.8	212.83	269.9	257.2	212.83	256.5	505	6570
	58.40*	C-95	0.595	8.435	8.279	10.625	10.125			8870	1604
	87.0		15.11	214.2	210.30	269.9	257.2			612	7135
	40.00*	MW-C-95	0.395	8.835	8.679	10.625	10.125	8.599	10.100	4230	1088
	59.6		10.03	224.4	220.45	269.9	257.2	218.41	256.5	292	4840
	43.50*	MW-C-95	0.435	8.755	8.599	10.625	10.125	8.599	10.100	5600	1193
	64.8		11.05	222.4	218.41	269.9	257.2	218.41	256.5	386	5307
	47.00*	MW-C-95	0.472	8.681	8.525	10.625	10.125	8.525	10.100	7100	1289
	70.0		11.99	220.5	216.54	269.9	257.2	216.54	256.5	489	5734
	53.50*	MW-C-95	0.545	8.535	8.379	10.625	10.125	8.379	10.100	8850	1477
	79.7		13.84	216.8	212.83	269.9	257.2	212.83	256.5	610	6570
	58.40*	MW-C-95	0.595	8.435	8.279	10.625	10.125			9860	1604
	87.0		15.11	214.2	210.30	269.9	257.2			686	7135
	40.00	P-110	0.395	8.835	8.679	10.625	10.125	8.599	10.100	3480	1260
	59.6		10.03	224.4	220.45	269.9	257.2	218.41	256.5	240	5605
	43.50	P-110	0.435	8.755	8.599	10.625	10.125	8.599	10.100	4430	1381
	64.8		11.05	222.4	218.41	269.9	257.2	218.41	256.5	306	6143
	47.00	P-110	0.472	8.681	8.525	10.625	10.125	8.525	10.100	5310	1493
	70.0		11.99	220.5	216.54	269.9	257.2	216.54	256.5	366	6641
	53.50	P-110	0.545	8.535	8.379	10.625	10.125	8.379	10.100	7930	1710
	79.7		13.84	216.8	212.83	269.9	257.2	212.83	256.5	547	7606

Table TD-2 *Continued*

1	2	3	4	5	6	7	8	9	10	11	12
Size: outside diameter D in mm	Nominal weight, threads and coupling lb/ft kg/m	Grade	Pipe		Threaded and coupled			Extreme-line		Collapse resistance P_c psi bar	Pipe body yield strength P_1 1000 lb kN
			Wall thickness t	Inside diameter d	Drift diameter	Regular	Special clearance	Drift diameter	Outside diameter W**		
						Outside diameter					
						W	W_c				
						in mm					
9⅝ 244.5	58.40* 87.0	P-110	0.595 15.11	8.435 214.2	8.279 210.30	10.625 269.9	10.125 257.2			9760 672	1957 8260
	40.00* 59.6	MW-125	0.395 10.03	8.835 224.4	8.679 220.45	10.625 269.9	10.125 257.2	8.500 216.41	10.100 256.5	3630 243	1432 6370
	43.50* 64.8	MW-125	0.435 11.05	8.755 222.4	8.599 218.41	10.625 269.9	10.125 257.2	8.500 216.41	10.100 256.5	4620 319	1578 6984
	47.00* 70.0	MW-125	0.472 11.99	8.681 220.5	8.525 216.54	10.625 269.9	10.125 257.2	8.525 216.54	10.100 256.5	5630 388	1697 7549
	53.50* 79.7	MW-125	0.545 13.84	8.535 216.8	8.379 212.83	10.625 269.9	10.125 257.2	8.379 212.83	10.100 256.5	8440 582	1943 8643
	58.40* 87.0	MW-125	0.595 15.11	8.435 214.2	8.279 210.30	10.625 269.9	10.125 257.2			10 550 727	2119 9386
	40.00* 59.6	MW-140	0.395 10.03	8.835 224.4	8.679 220.45	10.625 269.9	10.125 257.2	8.500 216.41	10.100 256.5	3630 243	1604 7135
	43.50* 64.8	MW-140	0.435 11.05	8.755 222.4	8.599 218.41	10.625 269.9	10.125 257.2	8.500 216.41	10.100 256.5	4730 326	1768 7820
	47.00* 70.0	MW-140	0.472 11.99	8.681 220.5	8.525 216.54	10.625 269.9	10.125 257.2	8.525 216.54	10.100 256.5	5660 406	1900 8452
	53.50* 79.7	MW-140	0.545 13.84	8.535 216.8	8.379 212.83	10.625 269.9	10.125 257.2	8.379 212.83	10.100 256.5	8700 606	2177 9684
	58.40* 87.0	MW-140	0.595 15.11	8.435 214.2	8.279 210.30	10.625 269.9	10.125 257.2			11 190 772	2363 10 511
	40.00* 59.6	V-150	0.395 10.03	8.835 224.4	8.679 220.45	10.625 269.9	10.125 257.2	8.500 216.41	10.100 256.5	3630 243	1716 7642
	43.50* 64.8	V-150	0.435 11.05	8.755 222.4	8.599 218.41	10.625 269.9	10.125 257.2	8.500 216.41	10.100 256.5	4756 326	1884 8380
	47.00* 70.0	V-150	0.472 11.99	8.681 220.5	8.525 216.54	10.625 269.9	10.125 257.2	8.525 216.54	10.100 256.5	6020 415	2038 9057
	53.50* 79.7	V-150	0.545 13.84	8.535 216.8	8.379 212.83	10.625 269.9	10.125 257.2	8.379 212.83	10.100 256.5	8976 618	2352 10 373
	58.40* 87.0	V-150	0.595 15.11	8.435 214.2	8.279 210.30	10.625 269.9	10.125 257.2			11 570 798	2532 11 263
	40.00* 59.6	MW-155	0.395 10.03	8.835 224.4	8.679 220.45	10.625 269.9	10.125 257.2	8.500 216.41	10.100 256.5	3630 243	1775 7895
	43.50* 64.8	MW-155	0.435 11.05	8.755 222.4	8.599 218.41	10.625 269.9	10.125 257.2	8.500 216.41	10.100 256.5	4756 326	1947 8661
	47.00* 70.0	MW-155	0.472 11.99	8.681 220.5	8.525 216.54	10.625 269.9	10.125 257.2	8.525 216.54	10.100 256.5	6056 417	2104 9359
	53.50* 79.7	MW-155	0.545 13.84	8.535 216.8	8.379 212.83	10.625 269.9	10.125 257.2	8.379 212.83	10.100 256.5	9026 622	2410 10 720
	58.40* 87.0	MW-155	0.595 15.11	8.435 214.2	8.279 210.30	10.625 269.9	10.125 257.2			11 720 808	2616 11 637

*Non-API.

Table TD-2 *Continued*

13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Internal yield pressure at minimum yield							Joint strength							
Plain-end or extreme-line	Round thread		Buttress/ BSDI Omega				Threaded and coupled						Extreme-line	
	Short	Long	Regular coupling		Special clearance coupling		Round thread		Buttress/ BSDI Omega					
							Short	Long	Regular coupling		Special clearance coupling			
			Same grade	Higher grade	Same grade	Higher grade			Same grade	Higher grade	Same grade	Higher grade	Standard joint	Optional joint
P_i psi bar									P_j 1000 lb kN					
3520	3520	3520	3520	3520	3520	3520	394	453	639	639	639	639		
243	243	243	243	243	243	243	1753	2015	2842	2842	2842	2842		
3950	3950	3950	3950	3950	3950	3950	462	520	714	714	714	714	770	770
272	272	272	272	272	252	272	2011	2313	3176	3176	3176	3176	3425	3425
3520	3520	3520	3520	3520	3520	3520	423	489	755	755	755	755		
243	243	243	243	243	243	243	1822	2175	3358	3358	3358	3358		
3950	3950	3950	3950	3950	3950	3950	486	561	843	843	843	843	975	975
272	272	272	272	272	252	272	2162	2495	3750	3750	3750	3750	4337	4337
5390		5390	5390	5390	4990	5390		694	926	926	926	926	975	975
372		372	372	372	344	372		3087	4119	4119	4119	4119	4337	4337
5930		5930	5930	5930	4990	5930		776	1016	1016	934	1016	975	975
409		409	409	409	344	409		3452	4519	4519	4155	4519	4337	4337
6440		6440	6440	6440	4990	6440		852	1098	1098	934	1098	1032	1032
444		444	444	444	344	444		3790	4884	4884	4155	4884	4591	4591
7430		7430	7430	7430	4990	7310		909	1257	1257	934	1229	1173	1053
512		512	512	512	344	504		4444	5591	5591	4155	5467	5218	4684
8110		8110	8110	8110	4990	7310		1098	1386	1386	934	1229		
559		559	559	559	344	504		4884	6072	6072	4155	5467		
5750		5750	5750	5750	5320	5750		727	947	947	934	947	975	975
396		396	396	396	367	396		3234	4212	4212	4155	4212	4337	4337
6330		6330	6330	6330	5320	6330		813	1038	1038	934	1038	975	975
436		436	436	436	367	436		3616	4617	4617	4155	4617	4337	4337
6870		6870	6870	6870	5320	6870		883	1122	1122	934	1122	1032	1032
474		474	474	474	367	474		3972	4991	4991	4155	4991	4591	4591
7930		7930	7930	7930	5320	7310		1047	1286	1286	934	1229	1173	1053
547		547	547	547	367	504		4657	5720	5720	4155	5467	5218	4684
8650		8650	8650	8650	5320	7310		1151	1396	1396	934	1229		
596		596	596	596	367	504		5120	6210	6210	4155	5467		
5750		5750	5750	5750	5320	5750		737	979	979	979	979	1027	1027
396		396	396	396	367	396		3278	4355	4355	4355	4355	4568	4568
6330		6330	6330	6330	5320	6330		825	1074	1074	963	1074	1027	1027
436		436	436	436	367	436		3670	4777	4777	4373	4777	4568	4568
6870		6870	6870	6870	5320	6870		906	1181	1181	963	1181	1096	1096
474		474	474	474	367	474		4026	5164	5164	4373	5164	4831	4831
7920		7930	7930	7930	5320	7310		1062	1328	1328	963	1229	1236	1109
547		547	547	547	367	504		4724	5912	5912	4373	5467	5494	4933
8650		8650	8650	8650	5320	7310		1167	1443	1443	963	1229		
596		596	596	596	367	504		5191	6419	6419	4373	5467		
6460		6460	6460	6460	5960	6460		804	1021	1021	963	1021	1027	1027
445		445	445	445	412	445		3576	4542	4542	4373	4542	4568	4568
7120		7120	7120	7120	5960	7120		899	1119	1119	963	1119	1027	1027
491		491	491	491	412	491		3999	4978	4978	4373	4978	4568	4568
7720		7720	7720	7720	5960	7310		957	1210	1210	963	1210	1087	1087
532		532	532	532	412	504		4390	5382	5382	4373	5382	4835	4835
8020		8020	8020	8020	5960	7310		1157	1396	1396	963	1229	1236	1109
615		615	615	615	412	504		5147	6165	6165	4373	5467	5494	4933
8740		8740	8740	8740	5960	7310		1272	1504	1504	963	1229		
672		672	672	672	412	504		5658	6690	6690	4373	5467		
8620		8620	8620	8620	6310	8620		847	1074	1074	1032	1074	1078	1078
470		470	470	470	435	470		3768	4777	4777	4591	4777	4795	4795
7510		7510	7510	7510	6310	7310		948	1178	1178	1032	1178	1078	1078
518		518	518	518	435	504		4217	5240	5240	4591	5240	4795	4795
8160		8160	8160	8160	6310	7310		1040	1273	1273	1032	1229	1141	1141
562		562	562	562	435	504		4626	5663	5663	4951	5467	5075	5075
9410		9410	9410	9410	6310	7310		1220	1488	1488	1032	1229	1297	1164
649		649	649	649	435	504		5427	6486	6488	4591	5467	5789	5178
10 280		10 280	10 280	10 280	6310	7310		1341	1683	1683	1032	1229		
709		709	709	709	435	504		5965	7042	7042	4591	5467		
8620		8620	8620	8620	6310	8620		868	1106	1106	1081	1106	1131	1131
470		470	470	470	435	470		3817	4920	4920	4809	4920	5031	5031
7510		7510	7510	7510	6310	7310		889	1213	1213	1081	1213	1131	1131
518		518	518	518	435	504		4266	5396	5396	4809	5396	5031	5031
8160		8160	8160	8160	6310	7310		1053	1311	1311	1081	1229	1167	1167
562		562	562	562	435	504		4684	5832	5832	4809	5467	5325	5325
9410		9410	9410	9410	6310	7310		1235	1502	1502	1081	1229	1359	1219
649		649	649	649	435	504		5494	6681	6681	4809	5467	6045	5422
10 280		10 280	10 280	10 280	6310	7310		1357	1650	1650	1081	1229		
709		709	709	709	435	504		6036	7251	7251	4809	5467		
7900		7900	7900	7900	7310	7900		998	1266	1266	1229	1266	1283	1283
545		545	545	545	504	545		4395	5631	5631	5467	5631	5707	5707
8700		8700	8700	8700	7310	8700		1166	1398	1398	1229	1398	1283	1283
600		600	600	600	504	600		4920	6174	6174	5467	6174	5707	5707
9440		9440	9440	9440	7310	9440		1213	1500	1500	1229	1500	1358	1358
651		651	651	651	504	651		5396	6672	6672	5467	6672	6041	6041
10 900		10 900	10 900	10 900	7310	9870		1422	1718	1718	1229	1573	1544	1398
752		752	752	752	504	667		6325	7642	7642	5467	6997	6868	6165

Table TD-2 *Continued*

13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Internal yield pressure at minimum yield							Joint strength							
Plain-end or extreme-line	Round thread		Buttress BSDI Omega				Threaded and coupled						Extreme-line	
	Short	Long	Regular coupling		Special clearance coupling		Round thread		Buttress BSDI Omega					
							Short	Long	Regular coupling		Special clearance coupling			
			Same grade	Higher grade	Same grade	Higher grade			Same grade	Higher grade	Same grade	Higher grade	Standard joint	Optional joint
			P_i psi bar						P_j 1000 lb kN					
11 900	11 900	11 900	11 900	7310	9870		1584	1985	1985	1229	1573			
820	820	820	820	504	687		6957	8296	8296	5467	6997			
8990	8990	8990	8990	8310	9990		1106	1383	1383	1327	1383	1386	1386	
619	619	619	619	573	619		4929	6196	6196	5903	6196	6174	6174	
8990	8990	8990	8990	8310	9990		1240	1527	1527	1327	1527	1386	1386	
682	682	682	682	573	682		5516	6792	6792	5903	6792	6174	6174	
10 730	10 730	10 730	10 730	8310	9870		1361	1650	1650	1327	1573	1499	1499	
740	740	740	740	573	687		6054	7340	7340	5903	6997	6534	6534	
12 390	12 390	12 390	12 390	8310	9870		1586	1890	1890	1327	1573	1687	1487	
854	854	854	854	573	687		7095	8407	8407	5903	6997	7415	6659	
13 520	13 520	13 520	13 520	8310	9870		1754	2052	2052	1327	1573			
932	932	932	932	573	687		7802	9128	9128	5903	6997			
10 060	10 060	10 060	10 060	9310	9870		1236	1552	1552	1476	1552	1542	1542	
693	693	693	693	642	687		5511	6904	6904	6561	6904	6859	6859	
11 070	11 070	11 070	11 070	9310	9870		1386	1702	1702	1476	1573	1542	1542	
763	763	763	763	642	687		6165	7571	7571	6561	6997	6859	6859	
12 010	12 010	12 010	12 010	9310	9870		1521	1839	1839	1476	1573	1632	1632	
828	828	828	828	642	687		6766	8180	8180	6561	6997	7259	7259	
13 670	13 670	13 670	13 670	9310	9870		1783	2107	2107	1476	1573	1863	1663	
956	956	956	956	642	687		7931	9372	9372	6561	6997	8243	7397	
15 150	15 150	15 150	15 150	9310	9870		1961	2297	2297	1476	1573			
1045	1045	1045	1045	642	687		8723	10 173	10 173	6561	6997			
10 770	10 770	10 770		9870			1326	1658		1573		1645	1645	
743	743	743		687			5896	7375		6997		7317	7317	
11 960	11 960	11 960		9870			1483	1818		1573		1645	1645	
818	818	818		687			6597	8087		6997		7317	7317	
12 670	12 670	12 670		9870			1628	1965		1573		1740	1740	
887	887	887		687			7242	8741		6997		7740	7740	
14 860	14 860	14 860		9870			1909	2236		1573		1976	1774	
1025	1025	1025		687			8492	9956		6997		8790	7891	
16 230	16 230	16 230		9870			2098	2444		1573				
1119	1119	1119		687			9332	10 871		6997				
11 130	11 130	11 130		10 300			1369	1712		1622		1696	1696	
767	767	767		710			6090	7615		7215		7544	7544	
12 260	12 260	12 260		10 300			1532	1877		1622		1696	1696	
845	845	845		710			6815	8349		7215		7544	7544	
13 300	13 300	13 300		10 300			1681	2028		1622		1796	1796	
917	917	917		710			7477	9021		7215		7985	7985	
15 360	15 360	15 360		10 300			1971	2323		1622		2038	1828	
1059	1059	1059		710			8767	10 333		7215		9065	8136	
16 770	16 770	16 770		10 300			2167	2522		1622				
1156	1156	1156		710			9639	11 218		7215				

Table TD-3: Properties of 7" casing

1	2	3	4	5	6	7	8	9	10	11	12
Size: outside diameter D in mm	Nominal weight, threads and coupling lb/ft kg/m	Grade	Pipe		Threaded and coupled			Extreme-line		Collapse resistance P_c psi bars	Pipe body yield strength P_y 1000 lb kN
			Wall thickness t	Inside diameter d	Drift diameter	Regular	Special clearance	Drift diameter	Outside diameter W		
						Outside diameter					
						W	W_c				
						in mm					
7 177.8	20.00 29.8	J-55	0.272 6.91	6.456 164.0	6.331 160.81	7.656 194.5				2270 157	318 1406
	23.00 34.3	J-55	0.317 8.05	6.366 161.7	6.241 158.52	7.656 194.5	7.375 187.3	6.151 156.24	7.390 187.7	3270 225	366 1626
	26.00 38.7	J-55	0.362 9.19	6.276 159.4	6.151 156.24	7.656 194.5	7.375 187.3	6.151 156.24	7.390 187.7	4320 298	415 1846
	29.00 29.8	K-55	0.272 6.91	6.456 164.0	6.331 160.81	7.656 194.5				2270 157	519 1406
	23.00 34.3	K-55	0.317 8.05	6.366 161.7	6.241 158.52	7.656 194.5	7.375 187.3	6.151 156.24	7.390 187.7	3270 225	366 1626
	26.00 38.7	K-55	0.362 9.19	6.276 159.4	6.151 156.24	7.656 194.5	7.375 187.3	6.151 156.24	7.390 187.7	4320 298	415 1846
	23.00 34.3	C-75	0.317 8.05	6.366 161.7	6.241 158.52	7.656 194.5	7.375 187.3	6.151 156.24	7.390 187.7	3770 260	488 2220
	26.00 38.7	C-75	0.362 9.19	6.276 159.4	6.151 156.24	7.656 194.5	7.375 187.3	6.151 156.24	7.390 187.7	5250 362	566 2516
	29.00 43.2	C-75	0.408 10.36	6.184 157.1	6.059 153.90	7.656 194.5	7.375 187.3	6.059 153.90	7.390 187.7	6790 466	634 2820
	32.00 47.7	C-75	0.453 11.51	6.094 154.8	5.969 151.61	7.656 194.5	7.375 187.3	5.969 151.61	7.390 187.7	8230 567	699 3109
	35.00 52.1	C-75	0.498 12.65	6.004 152.5	5.879 149.33	7.656 194.5	7.375 187.3	5.879 149.33	7.530 191.3	9710 669	763 3394
	38.00 56.6	C-75	0.540 13.72	5.920 150.4	5.795 147.19	7.656 194.5	7.375 187.3	5.795 147.19	7.530 191.3	10 680 736	822 3656
	23.00 34.3	L-80	0.317 8.05	6.366 161.7	6.241 158.52	7.656 194.5	7.375 187.3	6.151 156.24	7.390 187.7	3830 264	532 2366
	26.00 38.7	L-80	0.362 9.19	6.276 159.4	6.151 156.24	7.656 194.5	7.375 187.3	6.151 156.24	7.390 187.7	5410 373	604 2687
	29.00 43.2	L-80	0.408 10.36	6.184 157.1	6.059 153.90	7.656 194.5	7.375 187.3	6.059 153.90	7.390 187.7	7020 484	676 3007
	32.00 47.7	L-80	0.453 11.51	6.094 154.8	5.969 151.61	7.656 194.5	7.375 187.3	5.969 151.61	7.390 187.7	8600 593	745 3314
	35.00 52.1	L-80	0.498 12.65	6.004 152.5	5.879 149.33	7.656 194.5	7.375 187.3	5.879 149.33	7.530 191.3	10 180 702	814 3621
	38.00 56.6	L-80	0.540 13.72	5.920 150.4	5.795 147.19	7.656 194.5	7.375 187.3	5.795 147.19	7.530 191.3	11 390 785	877 3901
	23.00 34.3	N-80	0.317 8.05	6.366 161.7	6.241 158.52	7.656 194.5	7.375 187.3	6.151 156.24	7.390 187.7	3830 264	532 2366
	26.00 38.7	N-80	0.362 9.19	6.276 159.4	6.151 156.24	7.656 194.5	7.375 187.3	6.151 156.24	7.390 187.7	5410 373	604 2687
	29.00 43.2	N-80	0.408 10.36	6.184 157.1	6.059 153.90	7.656 194.5	7.375 187.3	6.059 153.90	7.390 187.7	7020 484	676 3007
	32.00 47.7	N-80	0.453 11.51	6.094 154.8	5.969 151.61	7.656 194.5	7.375 187.3	5.969 151.61	7.390 187.7	8600 593	745 3314
	35.00 52.1	N-80	0.498 12.65	6.004 152.5	5.879 149.33	7.656 194.5	7.375 187.3	5.879 149.33	7.530 191.3	10 180 702	814 3621
	38.00 56.6	N-80	0.540 13.72	5.920 150.4	5.795 147.19	7.656 194.5	7.375 187.3	5.795 147.19	7.530 191.3	11 390 785	877 3901
7 177.8	23.00* 34.3	MW-C-90	0.317 8.05	6.366 161.7	6.241 158.52	7.656 194.5	7.375 187.3	6.151 156.24	7.390 187.7	4030 278	599 2665
	26.00* 38.7	MW-C-90	0.362 9.19	6.276 159.4	6.151 156.24	7.656 194.5	7.375 187.3	6.151 156.24	7.390 187.7	5740 396	679 3020
	29.00* 43.2	MW-C-90	0.408 10.36	6.184 157.1	6.059 153.90	7.656 194.5	7.375 187.3	6.059 153.90	7.390 187.7	7560 523	759 3381
	32.00* 47.7	MW-C-90	0.453 11.51	6.094 154.8	5.969 151.61	7.656 194.5	7.375 187.3	5.969 151.61	7.390 187.7	9370 646	839 3732
	35.00* 52.1	MW-C-90	0.498 12.65	6.004 152.5	5.879 149.33	7.656 194.5	7.375 187.3	5.879 149.33	7.530 191.3	11 170 770	918 4075
	38.00* 56.6	MW-C-90	0.540 13.72	5.920 150.4	5.795 147.19	7.656 194.5	7.375 187.3	5.795 147.19	7.530 191.3	12 810 883	986 4386
	23.00 34.3	C-95	0.317 8.05	6.366 161.7	6.241 158.52	7.656 194.5	7.375 187.3	6.151 156.24	7.390 187.7	4180 286	632 2811
	26.00 38.7	C-95	0.362 9.19	6.276 159.4	6.151 156.24	7.656 194.5	7.375 187.3	6.151 156.24	7.390 187.7	5870 405	717 3189
	29.00 43.2	C-95	0.408 10.36	6.184 157.1	6.059 153.90	7.656 194.5	7.375 187.3	6.059 153.90	7.390 187.7	7830 539	803 3572
	32.00 47.7	C-95	0.453 11.51	6.094 154.8	5.969 151.61	7.656 194.5	7.375 187.3	5.969 151.61	7.390 187.7	9730 671	886 3937
	35.00 52.1	C-95	0.498 12.65	6.004 152.5	5.879 149.33	7.656 194.5	7.375 187.3	5.879 149.33	7.530 191.3	11 640 803	966 4297
	38.00 56.6	C-95	0.540 13.72	5.920 150.4	5.795 147.19	7.656 194.5	7.375 187.3	5.795 147.19	7.530 191.3	13 420 925	1041 4631
	23.00* 34.3	MW-C-95	0.317 8.05	6.366 161.7	6.241 158.52	7.656 194.5	7.375 187.3	6.151 156.24	7.390 187.7	5650 390	632 2811
	26.00* 38.7	MW-C-95	0.362 9.19	6.276 159.4	6.151 156.24	7.656 194.5	7.375 187.3	6.151 156.24	7.390 187.7	7990 536	717 3189

Table TD-3 *Continued*

1	2	3	4	5	6	7	8	9	10	11	12
Size: outside diameter D in mm	*Nominal weight, threads and coupling lb/ft kg/m*	*Grade*	*Pipe*		*Threaded and coupled*			*Extreme-line*		*Collapse resistance P_c psi bars*	*Pipe body yield strength P_y 1000 lb kN*
			Wall thickness t	*Inside diameter d*	*Drift diameter*	*Regular*	*Special clearance*	*Drift diameter*	*Outside diameter W*		
						Outside diameter					
						W	*W_c*				
						in mm					
	29.00* 43.2	MW-C-95	0.408 10.36	6.184 157.1	6.059 153.90	7.656 194.5	7.375 187.3	6.059 153.90	7.390 187.7	9200 634	803 3572
	32.00* 47.7	MW-C-95	0.453 11.51	6.094 154.8	5.969 151.61	7.656 194.5	7.375 187.3	5.969 151.61	7.390 187.7	10 400 718	885 3937
	35.00* 52.1	MW-C-95	0.498 12.65	6.004 152.5	5.879 149.33	7.656 194.5	7.375 187.3	5.879 149.33	7.530 191.3	11 640 803	966 4297
	38.00* 56.6	MW-C-95	0.540 13.72	5.920 150.4	5.795 147.19	7.656 194.5	7.375 187.3	5.795 147.19	7.530 191.3	13 420 925	1041 4631
	23.00* 34.3	P-110	0.317 8.05	6.366 161.7	6.241 158.52	7.656 194.5	7.375 187.3	6.151 156.24	7.390 187.7	4450 307	732 3256
	26.00 38.7	P-110	0.362 9.19	6.276 159.4	6.151 156.24	7.656 194.5	7.375 187.3	6.151 156.24	7.390 187.7	6210 428	830 3692
	29.00 43.2	P-110	0.408 10.36	6.184 157.1	6.059 153.90	7.656 194.5	7.375 187.3	6.059 153.90	7.390 187.7	8510 587	929 4132
	32.00 47.7	P-110	0.453 11.51	6.094 154.8	5.969 151.61	7.656 194.5	7.375 187.3	5.969 151.61	7.390 187.7	10 760 742	1025 4559
	35.00 52.1	P-110	0.498 12.65	6.004 152.5	5.879 149.33	7.656 194.5	7.375 187.3	5.879 149.33	7.530 191.3	13 010 897	1119 4978
	38.00 56.6	P-110	0.540 13.72	5.920 150.4	5.795 147.19	7.656 194.5	7.375 187.3	5.795 147.19	7.530 191.3	15 110 1042	1205 5360
7 177.8	23.00* 34.3	MW-125	0.317 8.05	6.366 161.7	6.241 158.52	7.656 194.5	7.375 187.3	6.151 156.24	7.390 187.7	4650 321	832 3701
	26.00* 38.7	MW-125	0.362 9.19	6.276 159.4	6.151 156.24	7.656 194.5	7.375 187.3	6.151 156.24	7.390 187.7	6480 445	944 4199
	29.00* 43.2	MW-125	0.408 10.36	6.184 157.1	6.059 153.90	7.656 194.5	7.375 187.3	6.059 153.90	7.390 187.7	8120 629	1056 4697
	32.00* 47.7	MW-125	0.453 11.51	6.094 154.8	5.969 151.61	7.656 194.5	7.375 187.3	5.969 151.61	7.390 187.7	11 720 808	1165 5182
	35.00* 52.1	MW-125	0.498 12.65	6.004 152.5	5.879 149.33	7.656 194.5	7.375 187.3	5.879 149.33	7.530 191.3	14 330 988	1272 5658
	38.00* 56.6	MW-125	0.540 13.72	5.920 150.4	5.795 147.19	7.656 194.5	7.375 187.3	5.795 147.19	7.530 191.3	16 760 1156	1370 6094
	23.00* 34.3	MW-140	0.317 8.05	6.366 161.7	6.241 158.52	7.656 194.5	7.375 187.3	6.151 156.24	7.390 187.7	4760 328	932 4146
	26.00* 38.7	MW-140	0.362 9.19	6.276 159.4	6.151 156.24	7.656 194.5	7.375 187.3	6.151 156.24	7.390 187.7	6690 461	1057 4702
	29.00* 43.2	MW-140	0.408 10.36	6.184 157.1	6.059 153.90	7.656 194.5	7.375 187.3	6.059 153.90	7.390 187.7	9500 659	1183 5262
	32.00* 47.7	MW-140	0.453 11.51	6.094 154.8	5.969 151.61	7.656 194.5	7.375 187.3	5.969 151.61	7.390 187.7	12 520 863	1304 5800
	35.00* 52.1	MW-140	0.498 12.65	6.004 152.5	5.879 149.33	7.656 194.5	7.375 187.3	5.879 149.33	7.530 191.3	15 490 1068	1424 6334
	38.00* 56.6	MW-140	0.540 13.72	5.920 150.4	5.795 147.19	7.656 194.5	7.375 187.3	5.795 147.19	7.530 191.3	18 260 1259	1534 6824
	23.00* 34.3	V-150	0.317 8.05	6.366 161.7	6.241 158.52	7.656 194.5	7.375 187.3	6.151 156.24	7.390 187.7	4800 331	896 4439
	26.00* 38.7	V-150	0.362 9.19	6.276 159.4	6.151 156.24	7.656 194.5	7.375 187.3	6.151 156.24	7.390 187.7	6690 475	1132 5035
	29.00* 43.2	V-150	0.408 10.36	6.184 157.1	6.059 153.90	7.656 194.5	7.375 187.3	6.059 153.90	7.390 187.7	9660 674	1267 5636
	32.00* 47.7	V-150	0.453 11.51	6.094 154.8	5.969 151.61	7.656 194.5	7.375 187.3	5.969 151.61	7.390 187.7	13 020 896	1398 6219
	35.00* 52.1	V-150	0.498 12.65	6.004 152.5	5.879 149.33	7.656 194.5	7.375 187.3	5.879 149.33	7.530 191.3	16 230 1119	1526 6788
	38.00* 56.6	V-150	0.540 13.72	5.920 150.4	5.795 147.19	7.656 194.5	7.375 187.3	5.795 147.19	7.530 191.3	19 240 1327	1644 7313
	23.00* 34.3	MW-155	0.317 8.05	6.366 161.7	6.241 158.52	7.656 194.5	7.375 187.3	6.151 156.24	7.390 187.7	4780 330	1032 4591
	26.00* 38.7	MW-155	0.362 9.19	6.276 159.4	6.151 156.24	7.656 194.5	7.375 187.3	6.151 156.24	7.390 187.7	6860 480	1179 5204
	29.00* 43.2	MW-155	0.408 10.36	6.184 157.1	6.059 153.90	7.656 194.5	7.375 187.3	6.059 153.90	7.390 187.7	9660 682	1319 5827
	32.00* 47.7	MW-155	0.453 11.51	6.094 154.8	5.969 151.61	7.656 194.5	7.375 187.3	5.969 151.61	7.390 187.7	13 220 912	1444 6423
	35.00* 52.1	MW-155	0.498 12.65	6.004 152.5	5.879 149.33	7.656 194.5	7.375 187.3	5.879 149.33	7.530 191.3	16 570 1142	1577 7015
	38.00* 56.6	MW-155	0.540 13.72	5.920 150.4	5.795 147.19	7.656 194.5	7.375 187.3	5.795 147.19	7.530 191.3	19 700 1366	1697 7549

* Non-API.

Table TD-3 *Continued*

13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Internal yield pressure at minimum yield							Joint strength							
Plain-end or extreme-line	Round thread		Buttress/ BSDI Omega				Threaded and coupled						Extreme-line	
	Short	Long	Regular coupling		Special clearance coupling		Round thread		Buttress/ BSDI Omega					
							Short	Long	Regular coupling		Special clearance coupling			
			Same grade	Higher grade	Same grade	Higher grade			Same grade	Higher grade	Same grade	Higher grade	Standard joint	Optional joint
			P_i psi bar						P_i 1000 lb kN					
10 760	10 760	10 050	10 760	6810	7690		778	922	922	617	702	882	780	
742	742	693	742	470	544		3465	4101	4101	2745	3123	3923	3470	
11 830	10 970	10 050	11 640	6810	7690		866	964	1007	617	702	964	861	
816	756	693	803	470	544		3848	4288	4479	2745	3123	4377	3919	
12 826	10 970	10 050	11 640	6810	7690		944	964	1066	617	702	1061	861	
884	756	693	803	470	544		4199	4288	4826	2745	3123	4720	3919	
6720	6720	6720	6720	7690	6720		690	752	752	702	752	832	832	
601	601	601	601	544	601		2624	3345	3345	3123	3345	3701	3701	
8960	8960	8960	8960	7690	8960		853	853	853	702	853	944	944	
687	687	687	687	544	687		3063	3794	3794	3123	3794	3754	3754	
11 220	11 220	11 220	11 220	7690	10 760		797	955	955	702	866	902	896	
774	774	774	774	544	742		3545	4248	4248	3123	3995	4012	3941	
12 460	12 460	11 640	12 460	7690	10 760		897	1053	1053	702	866	1002	896	
859	859	803	859	544	742		3990	4684	4684	3123	3995	4457	3941	
13 700	12 700	11 640	13 700	7690	10 760		996	1096	1156	702	866	1116	1002	
945	876	803	945	544	742		4430	4875	5115	3123	3995	4973	4457	
14 660	12 700	11 640	14 660	7690	10 760		1087	1096	1239	702	866	1207	1002	
1024	876	803	1024	544	742		4835	4875	5511	3123	3995	5369	4457	
9010	9010	9010	9010	6970	9010		655	823	823	758	823	899	899	
683	683	683	683	618	683		2914	3661	3661	3372	3661	3999	3999	
11 310	11 310	11 310	11 310	6970	10 760		799	934	934	758	842	911	911	
780	780	780	780	618	742		3421	4155	4155	3372	3745	4052	4052	
12 760	12 760	12 760	12 760	6970	10 760		885	1045	1045	758	842	973	956	
879	879	879	879	618	742		3937	4648	4648	3372	3745	4284	4253	
14 160	14 160	13 220	14 160	6970	10 760		996	1152	1152	758	842	1082	957	
976	976	911	976	618	742		4430	5124	5124	3372	3745	4813	4257	
15 660	14 430	13 220	15 660	6970	10 760		1106	1163	1266	758	842	1207	1061	
1073	995	911	1073	618	742		4920	5262	5596	3372	3745	5369	4809	
16 660	14 430	13 220	15 870	6970	10 760		1207	1163	1315	758	842	1203	1061	
1164	995	911	1094	618	742		5369	5262	5849	3372	3745	5796	4809	
11 690	11 090	11 090	11 090	10 040	10 760		724	917	917	842	866	966	966	
765	765	765	765	692	742		3221	4079	4079	3745	3995	4439	4439	
12 670	12 670	12 670	12 670	10 040	10 760		856	1040	1040	842	866	1012	1012	
874	874	874	874	692	742		3803	4626	4626	3745	3995	4502	4502	
14 280	14 280	14 280	14 280	10 040	10 760		966	1164	1164	842	866	1061	1063	
985	985	985	985	692	742		4973	5178	5178	3745	3995	4809	4729	
15 990	15 990	14 510	15 870	10 040	10 760		1107	1283	1283	842	866	1202	1064	
1094	1094	1021	1094	692	742		4924	5707	5707	3745	3995	5347	4733	
17 430	15 170	14 510	15 870	10 040	10 760		1229	1315	1461	842	866	1341	1201	
1202	1115	1021	1094	692	742		5467	5849	6232	3745	3995	5965	5342	
15 900	15 170	14 510	15 870	10 040	10 760		1341	1315	1402	842	866	1447	1201	
1303	1115	1021	1094	692	742		5965	5849	6236	3745	3995	6437	5342	
11 990	11 990	11 990		10 760			778	979		866		1066	1066	
820	820	820		742			3452	4355		3995		4737	4737	
13 570	13 570	13 570		10 760			912	1110		866		1079	1079	
936	936	936		742			4057	4938		3995		4800	4800	
15 300	15 300	15 300		10 760			1046	1243		866		1183	1134	
1055	1055	1055		742			4666	5529		3995		5129	5044	
16 990	16 990	15 870		10 760			1160	1370		866		1282	1134	
1171	1171	1094		742			5240	6094		3995		5703	5044	
18 990	17 320	15 870		10 760			1310	1402		866		1431	1281	
1268	1194	1094		742			5827	6236		3995		6365	5698	
20 260	17 320	15 870		10 760			1430	1402		866		1544	1281	
1396	1194	1094		742			6261	6236		3995		6868	5698	
12 280	12 280	12 280		11 120			890	1010		926		1066	1066	
847	847	847		767			3559	4493		4119		4884	4884	
14 030	14 030	14 030		11120			940	1140		926		1113	1113	
967	967	967		767			4181	5098		4119		4951	4951	
15 810	15 810	15 810		11 120			1081	1262		926		1186	1186	
1090	1090	1090		767			4809	5703		4119		5289	5200	
17 550	17 550	16 400		11 120			1217	1414		926		1322	1178	
1210	1210	1131		767			5413	6290		4119		5881	5204	
18 300	17 990	16 400		11 120			1361	1446		926		1475	1321	
1331	1234	1131		767			6010	6432		4119		6561	5876	
20 690	17 990	16 400		11 120			1475	1446		926		1592	1321	
1443	1234	1131		767			6561	6432		4119		7082	5876	

Appendix E: PDF Data for Variables

E.1: Tubular property data

Variable	Distribution	Units	Mean	COV	Curtailment
Yield stress	Normal	Actual/nominal	Varies	0.035	None
UTS	Normal	Actual/nominal	Varies	0.025	None
Local wall thickness	Normal	Actual/nominal	1.000	0.030	None
Min wall thickness	Normal	Actual/nominal	0.960	0.025	None
Average wall thickness	Normal	Actual/nominal	1.000	0.015	None
Local outside diameter	Normal	Actual/nominal	1.005	0.002	None
Average outside diameter	Normal	Actual/nominal	1.005	0.001	None
Eccentricity	Log-normal	Actual (%)	7.500	0.510	None
Ovality	Log-normal	Actual (%)	0.250	0.560	None
Residual stress	Normal	Actual/yield	-0.230	0.240	None
Young's modulus	Normal	Actual/nominal	1.000	0.035	None
Poisson's ratio	Normal	Actual/nominal	1.000	0.025	None
Linear expansivity	Normal	Actual/nominal	1.000	0.040	None
Thermal degradation of yield	Log-normal	Actual (%/^{o}F)	0.035	0.500	0.090
Thermal degradation of UTS	Log-normal	Actual (%/^{o}F)	0.030	0.700	0.080

E.2: mean yield stress and UTS

Grade	Yield Stress actual/nominal	UTS actual/nominal
J55	1.090	1.400
K55	1.090	1.070
L80	1.090	1.070
N80	1.180	1.070
C90	1.060	1.070
C95	1.060	1.070
P110	1.080	1.070
Q125	1.060	1.070
V150	1.030	1.070

E.3: Load PDF data

Variable	Distribution (units)	Quantity	Exploration Well	Appraisal Well	Development Well
Pore pressure	Normal (actual/predicted)	Mean COV	1.00 0.11	1.00 0.06	1.00 0.06
Fracture gradient	Normal (actual/predicted)	Mean COV	1.01 0.11	1.01 0.11	1.00 0.06
Reservoir pressure	Normal (actual/predicted)	Mean COV	1.00 0.13	1.00 0.12	1.00 0.12
Reservoir temp	Normal (actual/predicted)	Mean COV	1.00 0.15	1.00 0.06	1.00 0.06
Production fluid density	Normal (actual/predicted)	Mean COV	1.00 0.18	- -	1.00 0.10
Dogleg severity DLS<3°/100ft	Rayleigh (actual, °/100ft)	Mean COV	3.0 n/a	3.0 n/a	3.0 n/a
Dogleg severity DLS>3°/100ft	Normal (actual/predicted)	Mean COV	(1) 0.3	(1) 0.3	(1) 0.3

Notes (1): mean DLS= target plus 2°/100 ft for low builds
target plus 4°/100ft for high builds

E.4: Model uncertainties

Variable	Distribution	Units	Mean	COV
Burst strength	Normal	Actual/predicted	1.000	0.047
Collapse strength	Normal	Actual/predicted	0.977	0.076
Top-of-bubble pressure during kick circulation	Normal	Actual/predicted	0.930	0.045
-at casing shoe (WBM)	Normal	Actual/predicted	0.930	0.120
-at wellhead (WBM)	Normal	Actual/predicted	0.860	0.045
-at casing shoe (OBM)	Normal	Actual/predicted	0.860	0.190
-at wellhead (OBM)				

E.5: Kick load PDF parameters

Variable	Distribution (units)	Curtailment	Quantity	Well Type		Well Duty		Installation Type	
				Explor-ation	Devel-opment	HPHT	Non-Critical	Floating	Fixed
Kick influx volume V -lower population	Weibull[(1)] (bbl)	None	Mean COV	2.0 1.0	15 1.0	20 1.0	15 1.0	20 1.0	15 1.0
-upper population	weibull[(2)] (bbl)	None	Mean COV	190 0.55	190 0.55	190 0.55	190 0.55	190 0.55	190 0.55
-ratio upper/lower	n/a	n/a	Lower Upper	0.80 0.20	0.90 0.10	0.80 0.20	0.90 0.10	0.80 0.20	0.90 0.10
Kick intensity -lower v, surface & intermediate	Weibull[(1)] (ppg)	6.0	Mean COV	1.1 1.0	1.1 1.0	1.1 1.0	1.1 1.0	1.1 1.0	1.1 1.0
-low V, production	Weibull[(1)] (ppg)	8.0	Mean COV	1.3 1.2	1.3 1.2	1.3 1.2	1.3 1.2	1.3 1.2	1.3 1.2
-upper V, all casing	Weibull[(1)] (ppg)	3.0	Mean COV	0.6 1.0	0.6 1.0	0.6 1.0	0.6 1.0	0.6 1.0	0.6 1.0

Notes:

1) Two parameter – scale parameter, a=0
2) Three parameter – scale parameter, a=70 bbl

Appendix F: Software introduction

As the final aim of the present research, QRACD software package has been developed. The software name QRACD comes from Quantitative Risk Assessment for Casing Design. The software was developed using MS Visual C++ with MFC (Microsoft Foundation Class). Monte Carlo Simulation and AIS Tail Model were currently built into the software.

The main subroutine, CACDLib (Computer Aided Casing Design Library), was programmed as a DLL (Dynamic Link Library) using FORTRAN and called by the user interface. It contains the main calculation part, MCS and AIS Tail Model. RANDLib is another DLL developed by FORTRAN. It provides a large range of general methods to produce random numbers. The detailed user menu about how to use CACDLib and RANDLib will be given in additional document to provide user the further developing information.

The software employs ActiveX Control. The main one is FORMULA 1, which was developed by other Internet programmer. It provides user the plot methods (see Figure F-7). Therefore, FORMULA 1 must be installed before installing QRACD. Further development need to be done to get rid of all the plug-ins so that the software can be completely independent.

A brief menu about how to use the current software is given as the following charts.

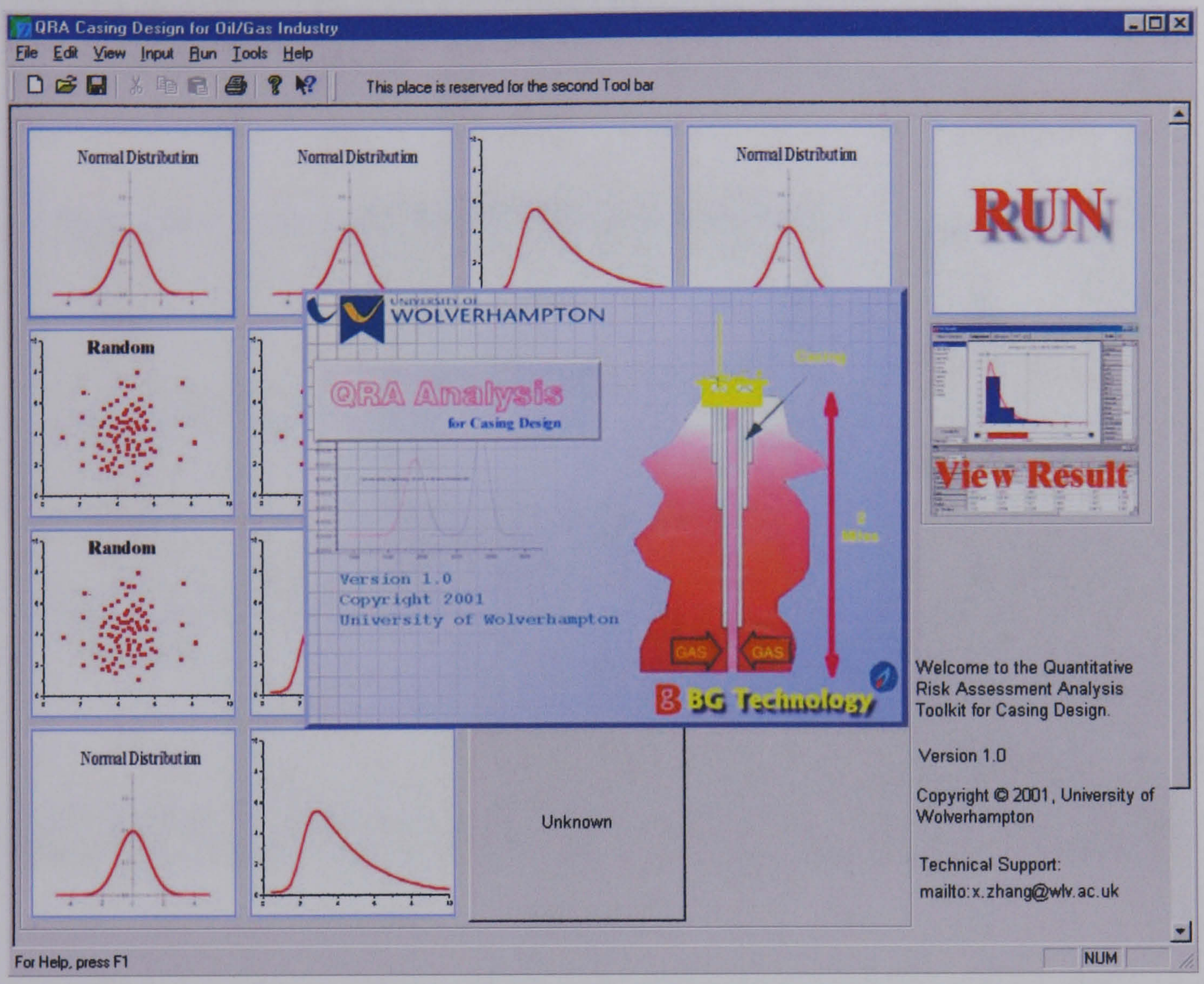


Figure F-1

Figure 1 shows the start flash screen. It provides version information. Further development will put user register information and user ID.

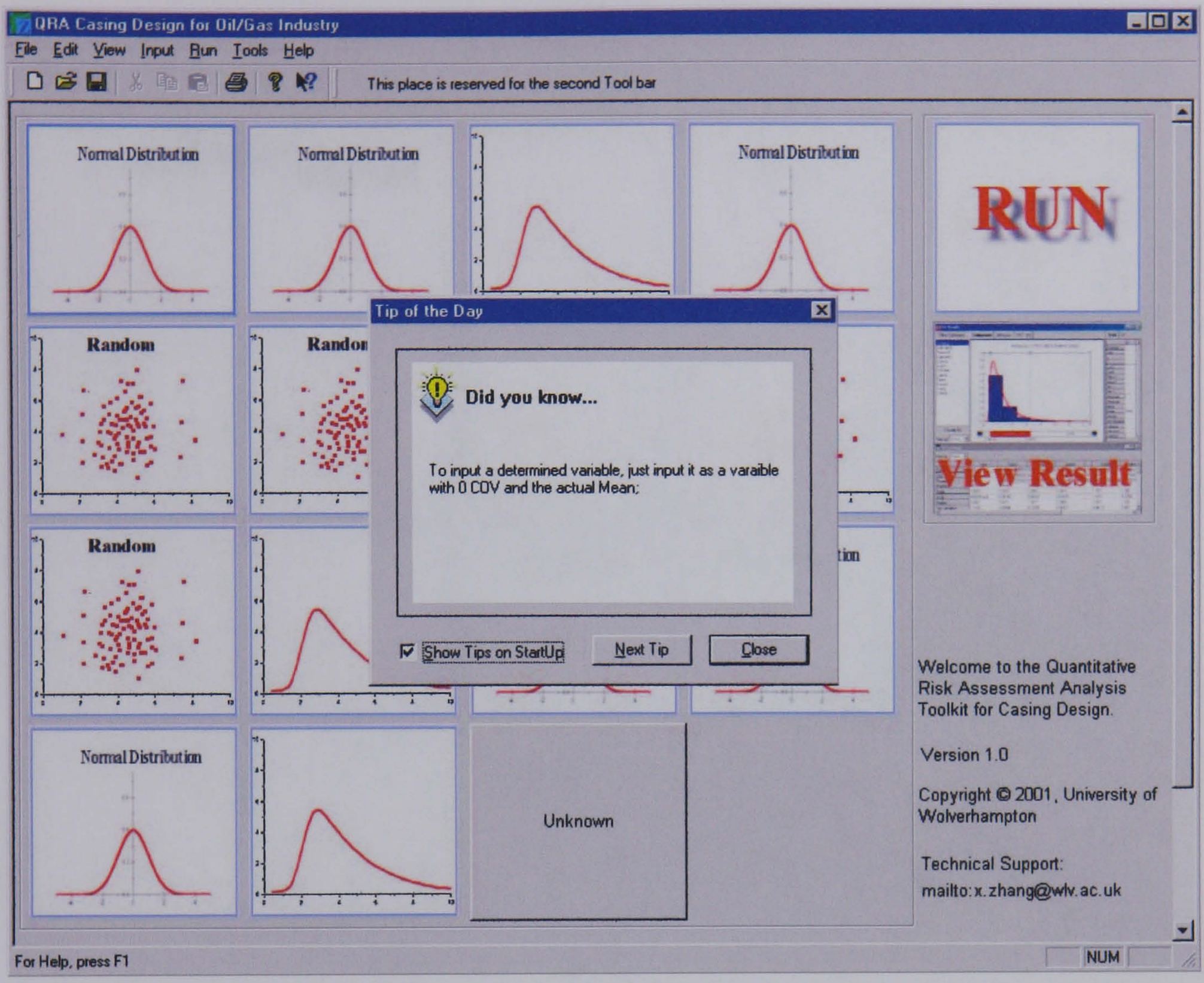


Figure F-2

Tips of the day will tell user some shortcut keys and basic information about casing design.

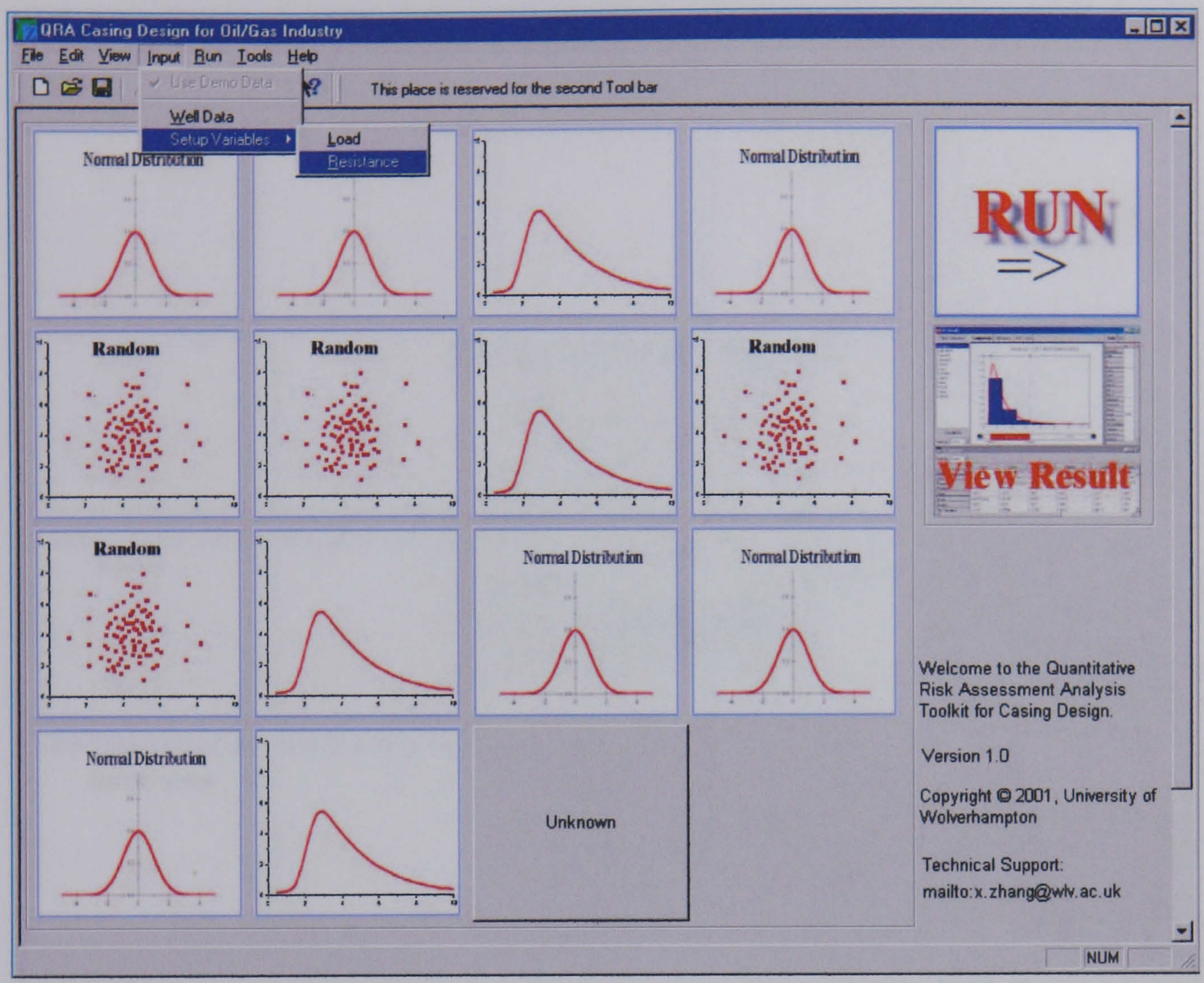


Figure F-3

Figure F-3 shows the main screen of the software, all variables involved in the design procedure are listed in the main window to give user a more direct view. The distribution types are plotted as graphics to give easy use. One thing that has to be stated is that the random "distribution" shown in the graph means that the variables are random distributed and the simulation will use the built-in dataset.

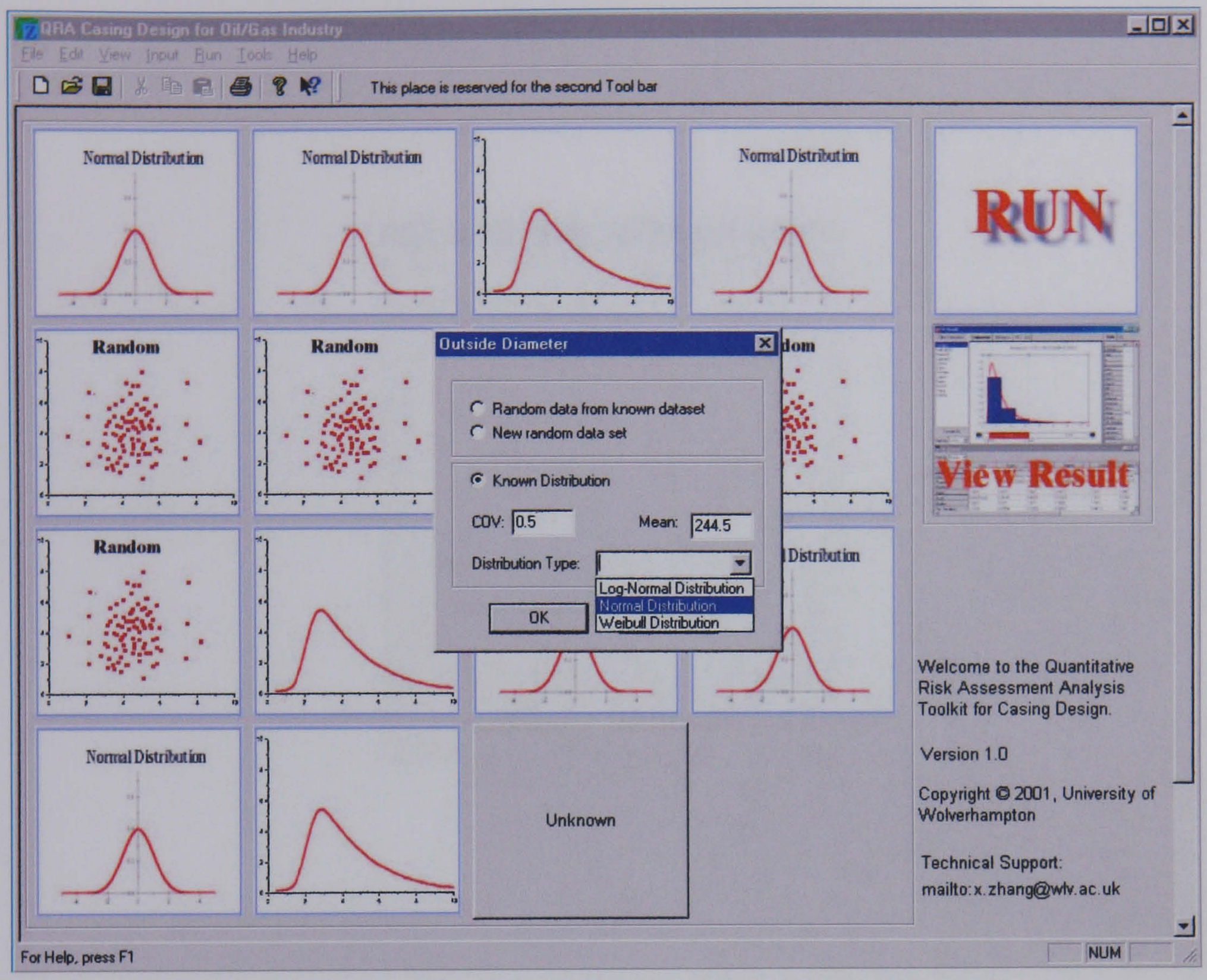


Figure F-4

Double click the variables which you want to input or change the distribution information of the variables. It will bring you a popup window, such as Outside diameter window in Figure F-4. If you select Random Data from known dataset, the software will initialise this variable using built-in database. Radio Button "New dataset" is not functional at this moment. It is supposed to add new random data or new variables in the further development.

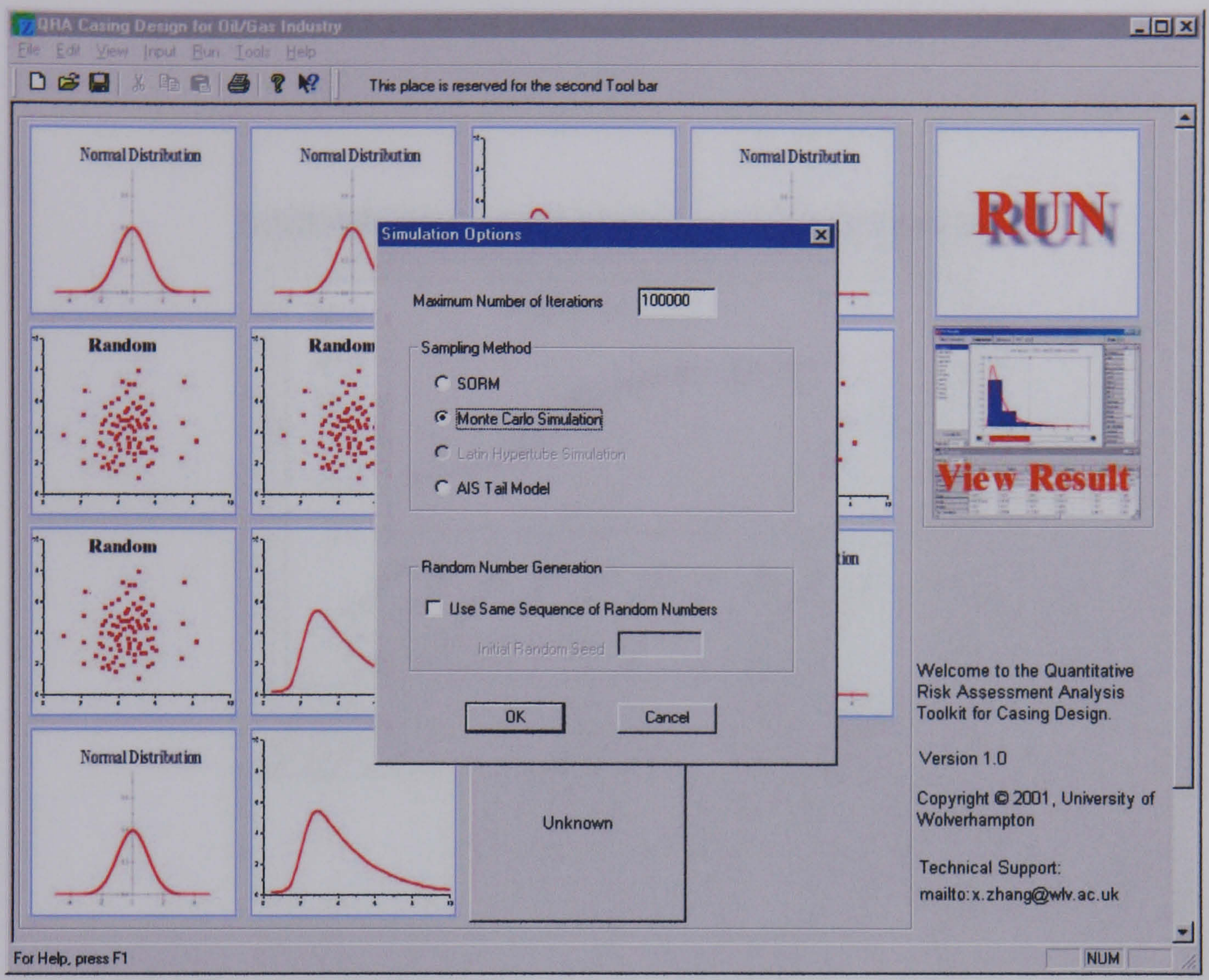


Figure F-5

After, setting up all the variables, go to "Tools" -> "Simulation Options", it will bring you a window as shown in Figure F-5. Here, you can set up the simulation methods you want to use. Note: FORM and Latin Hypercube method are not functional yet.

Then go to pull down menu "Run" and select "Collapse Analysis" or "Burst Analysis". Then double click button "RUN" on the up-right hand of the window. The simulation will be fulfilled on the backstage. The bigger numbers of iteration you input, the longer time the simulation will take.

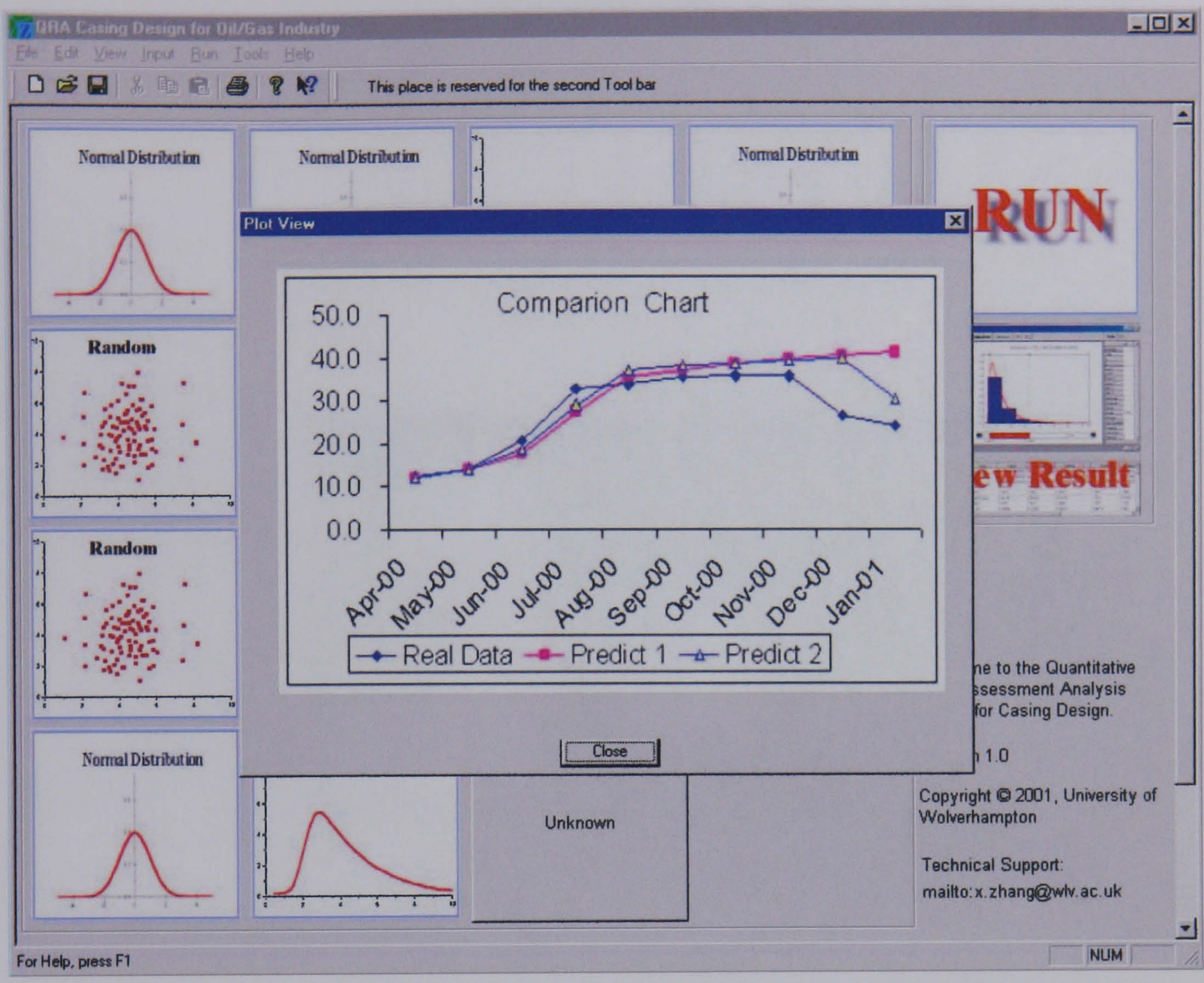


Figure F-6

After simulation, double click the button "View Result" bellow "RUN" button, a graphic window will bring to the front, which provides the plotview of the result. To define the plot type and plot data, you have to the next step.

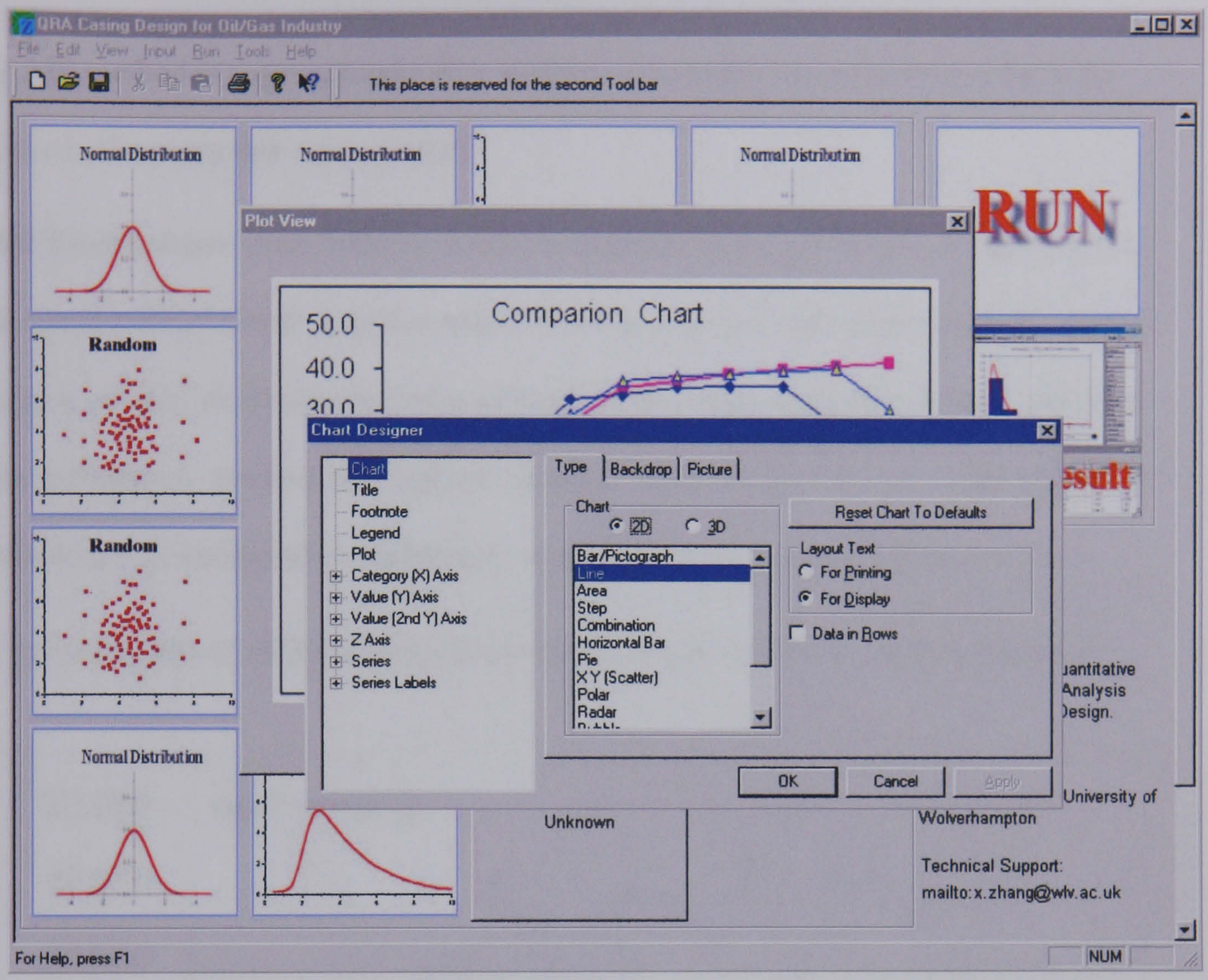


Figure F-7

Right click the plot window in Figure F-6 and select "Chart Designer" in the popup context menu, you will see Figure F-7. It provides you a very easy way to define the plot type and the data you want to use. (Same as the chart designer in Excel).

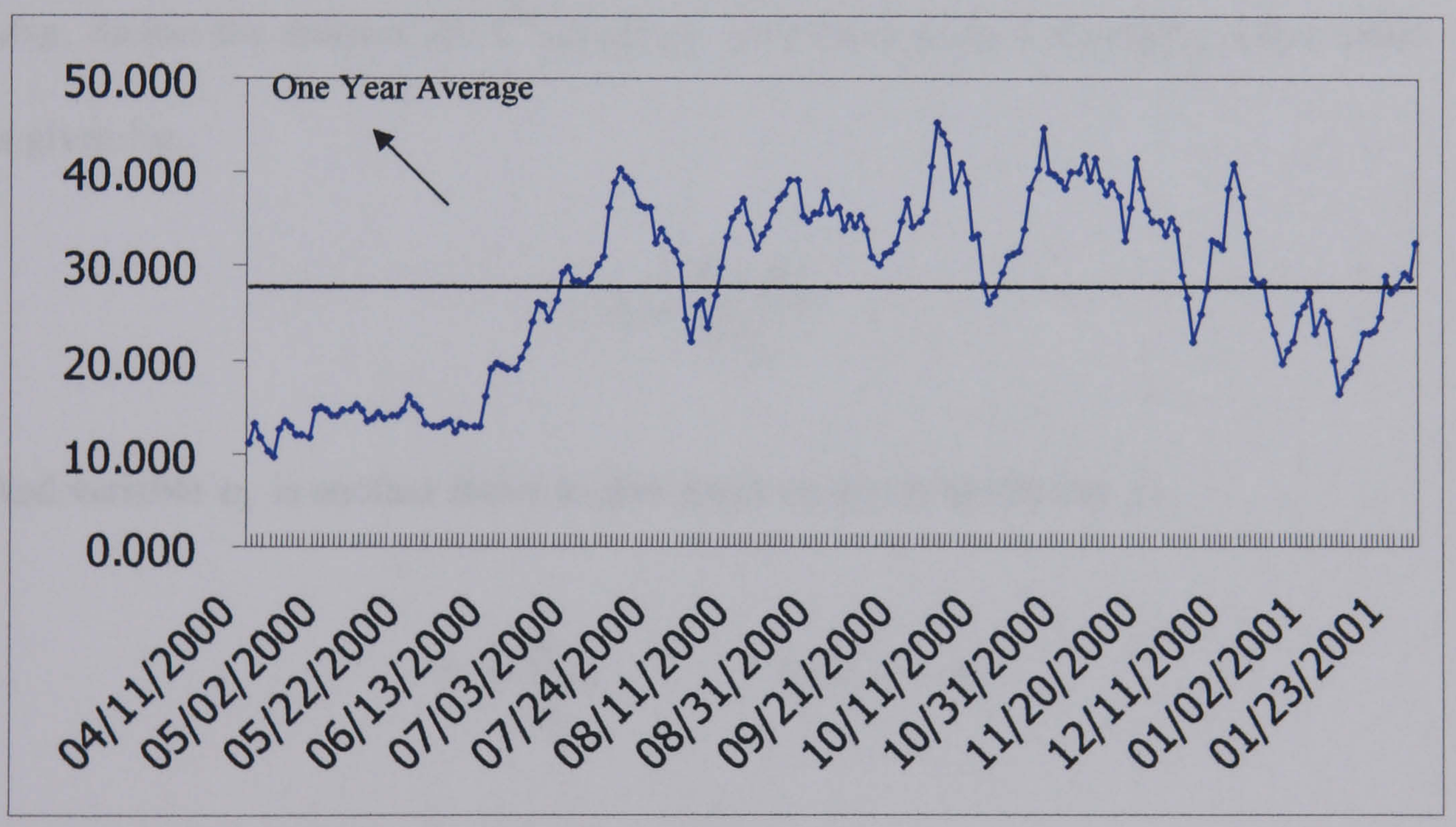
As mentioned earlier in this chapter, this function uses ActiveX Control plug-in: FORMULA 1, which should be pre-installed. Otherwise, the "View Result" button won't work.

Appendix G: Investigation of the generalisation of AIS model

As AIS method is a generalized QRA method, it is hereby using to analyse the share prices in the every day stock market.

DDI Corporation is an industry leader in manufacturing quick-turn, technologically advanced printed circuit board products. It offers services that enable its customers to shorten product development cycles and reduce their time to market for new products. The company's services are used to develop products for use in communications switching and transmission equipment, wireless base stations, work stations, etc.

A historical data of DDI Corporation (NASDA: DDIC) has been got from Internet.



The above figure shows the changing curve of the daily DDI share price from 11th Apr, 2000 to 26th Jan, 2001, together with the average line through the year.

The problems now is in how to build a suitable probability model to predict the changing direction of share price, based on the historical data, during the next

working day (or the next considered point). Here, we have to emphasis that the probability model we are going to build is base purely on the historical data. Actually, other important factors that affect the share prices include market condition, economic policy, etc. the model is just for exploring the feasibility of using AIS method to solve the problem.

Assume:

$$P_{n+1} = Avg_i[1 + (a_1 f_1 + a_2 f_2 + \cdots + a_n f_n)]$$

where the total historical data was divided into n groups, each group has the equal numbers of data. P_{n+1} is the price of next future point we are considering. Variable Avg_i denote the mean of the i^{th} group. f_i is introduced as a weighting factor which is given by:

$$f_i = \frac{P_i - P_{i-1}}{P_{i-1}}$$

And variable a_n is another factor to give more weight to the recent f_i,

$$a_i = \sum_{k=1}^{i-1} a_k \qquad i = 2,3,\cdots,n$$

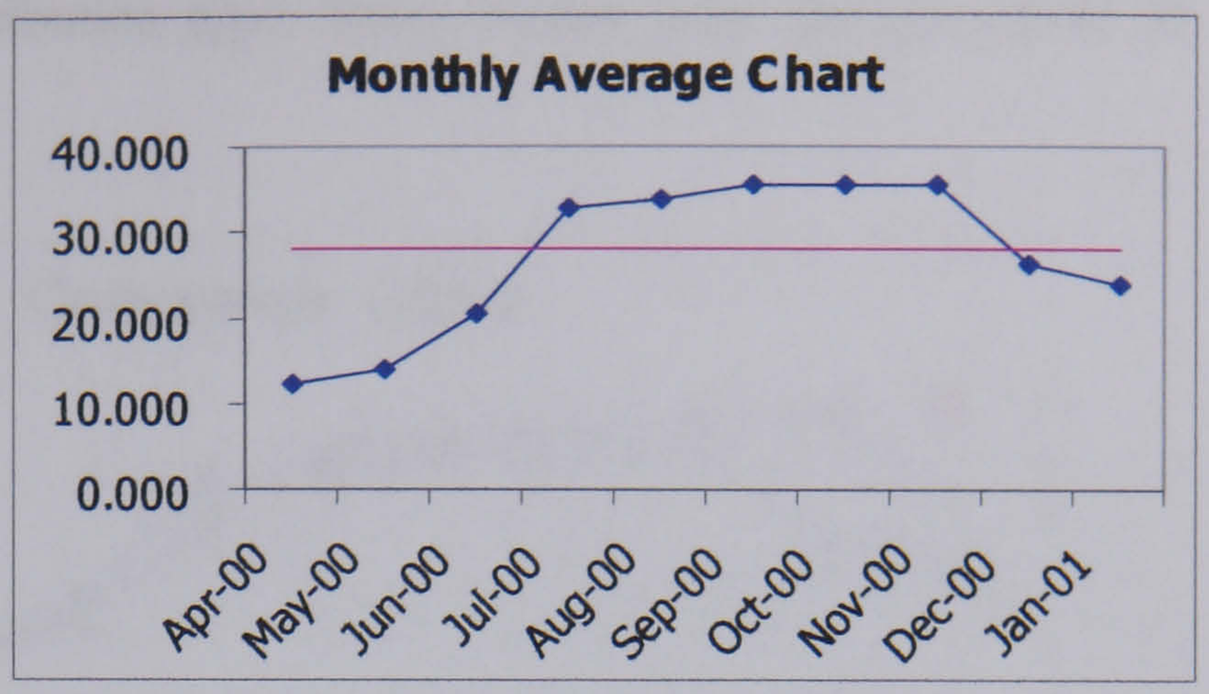
(During the simulation, we initialise $a_1 = 0.35$)

Thus the probability model can be written as,

$$g = P_{n+1} - P_n = g(Avg)$$

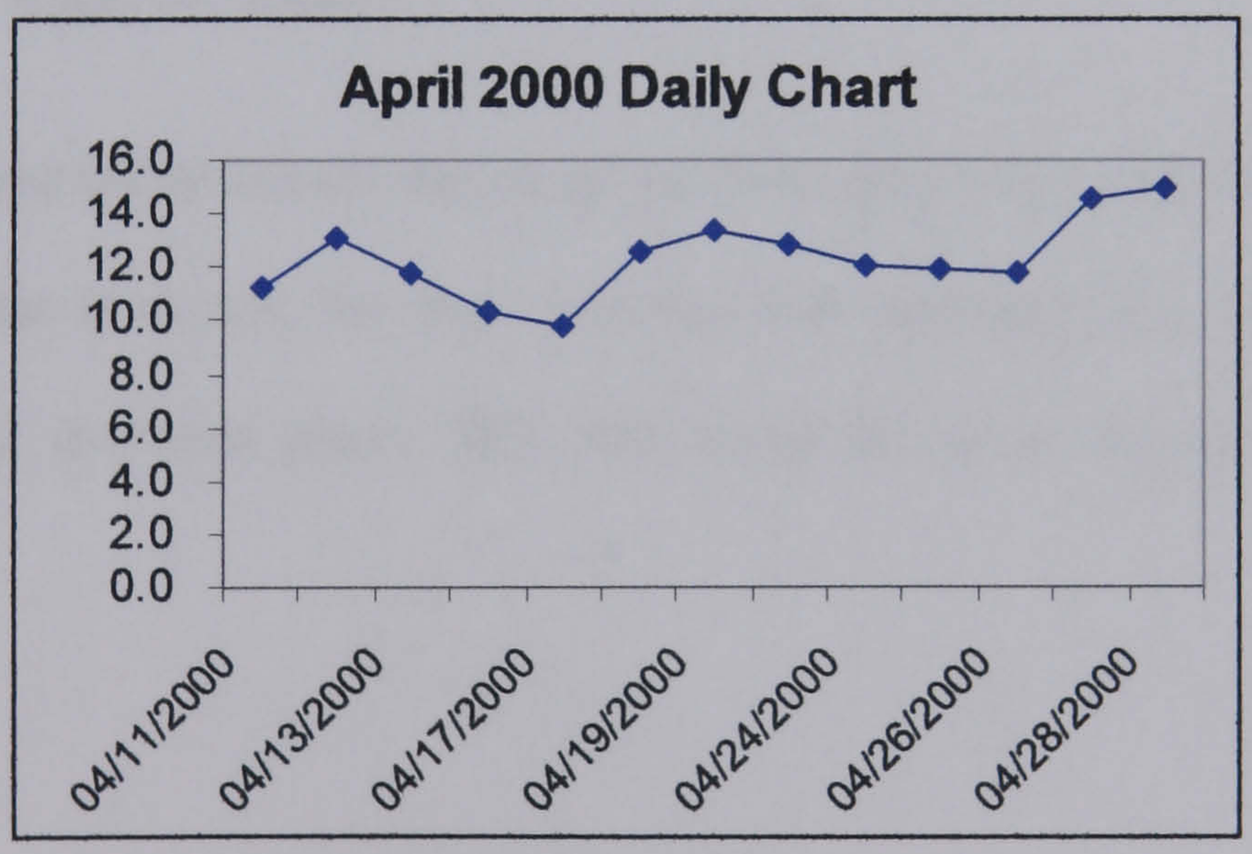
We can easily found that probability g is the function of unique variable Avg_i.

Monthly Average	
Apr-00	12.319
May-00	14.102
Jun-00	20.557
Jul-00	32.831
Aug-00	33.777
Sep-00	35.444
Oct-00	35.710
Nov-00	35.699
Dec-00	26.482
Jan-01	24.010

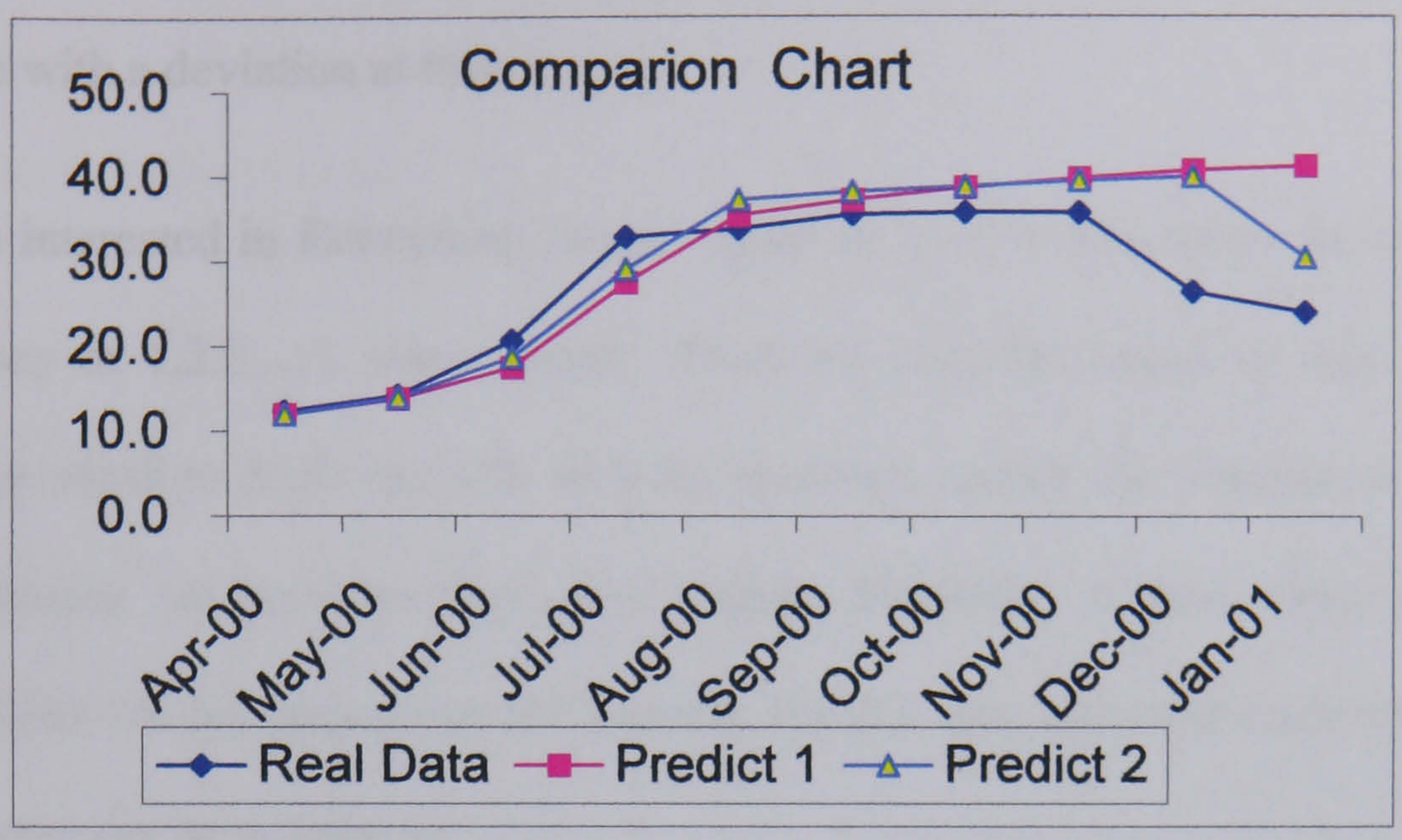


The above figure gives the monthly average chart. We are now trying to get a predicted curve of monthly chart.

Suppose we've known the average price of April 2000, 12.319, to predict the average of May 2000, as the probability model describes, we have to use the daily historical data of April (shown below).



It is obviously not a known distribution type. Thus, we use these thirteen points as



first samples of the simulation procedure. Run the simulation once, we got a predicted value 12.0, add this new value into the database and we will have 14 points in our new database, that is to say, the next sample database simulation will have 14 points which contents the new data. Repeating the simulation, the predicted curve is given in the above figure as "Predict 1".

In predict 2, suppose we've known the actual previous data every time before we run the simulation. That is to say, we don't use the new predicted data as one of the samples to predict the next point. This will avoid the error continuously being transferred.

A refined simulation model:

As described by Giorgio P Szego and Karl Shell (1972), a forecasting model is as following:

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_{t-1} x_0$$

where $\phi_i, (i = 0,1,2,\cdots,t-1)$ will be determined by the iteration. To determining it, the historical data has to be used to make a best fit. This equation gives a value estimation of share price with a deviation at time t.

We are often interested in forecasting future values of an observed series for several lead times, say at 1,2,3,...,L steps ahead. When we forecast values at lead times greater that or equal to 2 (l>=2) with an autoregressive model, the forecasted value will be dependent on previous forecasted values. However, if new observations become available we can update our old forecast. By this idea, our model described in the previous part can be written as:

$$P_{n+1} = P_{Hist}(\cdot) * (1 + \frac{1}{n}\sum_{j}^{n} \theta_j f_j)$$

$$P_{n+l} = P^{*}_{Hist}(\cdot) * (1 + \frac{1}{n}\sum_{j}^{n+l} \theta_j f_j)$$

where θ_j is given by

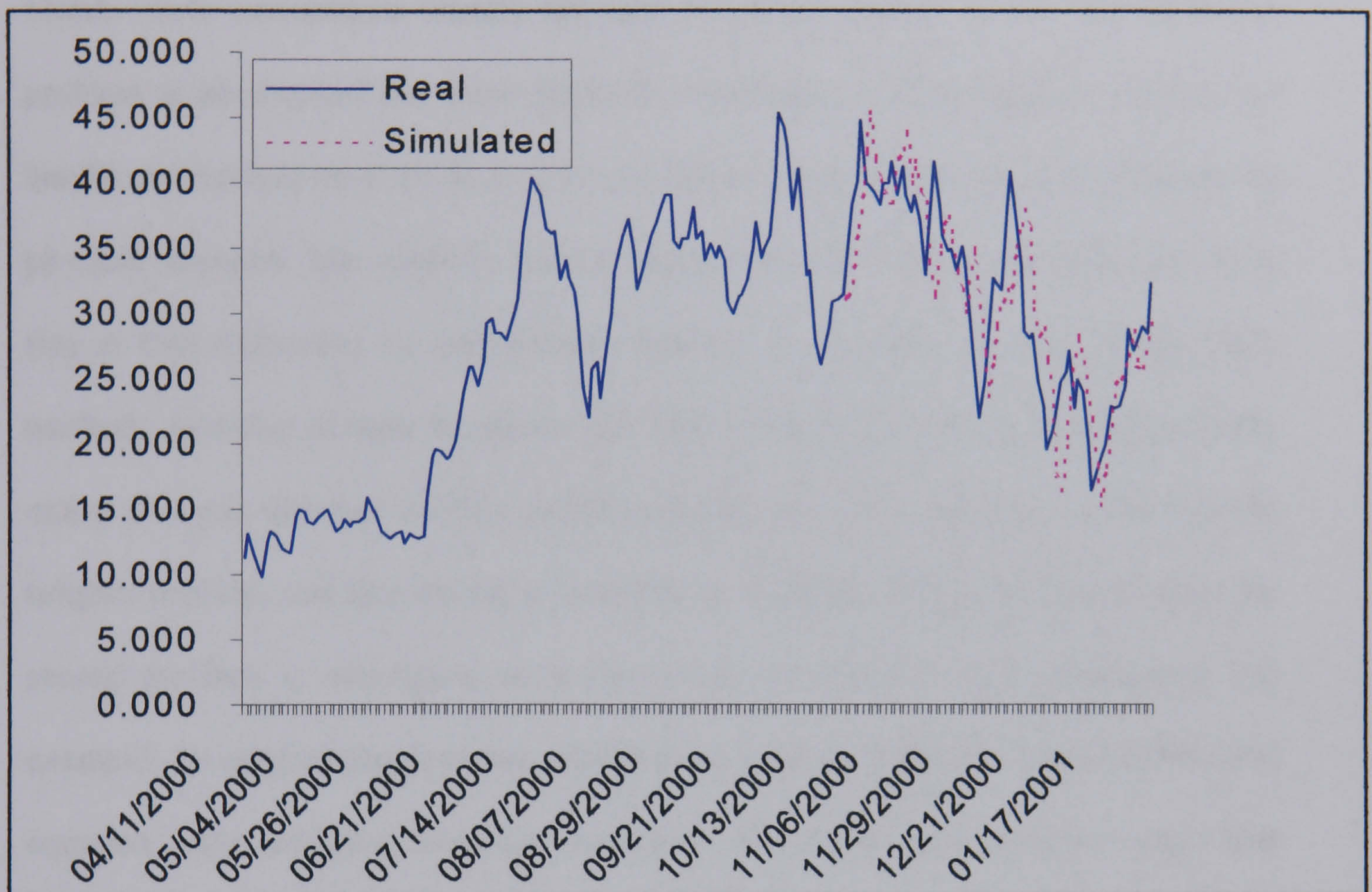
$$\begin{aligned}
&\theta_j = 0,\ j < 0 \\
&\theta_0 = 1, \\
&\theta_1 = \phi_1, \\
&\theta_2 = \phi_1\theta_1 + \phi_2 \\
&\vdots \\
&\theta_j = \phi_1\theta_{j-1} + \cdots + \phi_j\theta_0
\end{aligned}$$

During the simulation, $P_{Hist}(\cdot)$ is random distributed, the forecasting value P_{n+1} is therefor completely related to the historical data, while P_{n+l} relating to the old forecast. When the real value of P_{n+1} becomes available, we update our database and the forecast value $l-1$ steps ahead has to be recalculated.

The probability model, g is then a function of random distributed historical data,

$$g = P_{n+1} - P_n = g(P_{Hist})$$

Then the simulated result can be show in the following plot.



It can be seen from this example that the AIS method can be used to analyse share market, and give a consultative prediction value. Of course, the share prices are depending on lots of factors, instead of only the few factors considered in this trial, therefor, as a pure data analysis, this example is just for demonstrated that the AIS method can be used to this area.

This example also gives the further evidence to show that AIS method is a generalize method.

Appendix H: Monte Carlo Method for Solving Multivariable Problems

1 The Nature of Monte Carlo Methods

Monte Carlo calculations usually fall into two main classes. In the first class, the problem to be attacked has some physical probabilistic structure and, in essence, we handle our random numbers in such a way that they more or less directly simulate the physical situation. The second class of Monte Carlo calculations concerns problems that at first sight have no probabilistic features. To be able to apply Monte Carlo methods, however, it must be shown that such a first impression is misleading to the extent that can discover another problem having the same numerical answer as the original problem and also having a probabilistic structure. It then suffices to solve the second problem by simulation, as in the first class of Monte Carlo calculations. For example, the original problem may require the solution of Laplace's partial differential equation under assigned boundary conditions. The theory of probability states that certain properties of random walks also satisfy Laplace's equation under a wide variety of boundary conditions. Therefore a property must be addressed such that the boundary conditions are the same as those in the original assigned problem; then the behaviour of this property can be studied in a sample of random walks, each walk generated on a computer by a sequence of random numbers.

In either class of Monte Carlo calculation, the desired numerical results are identified, either directly by the statement of a physical situation or indirectly from the theory of probability, with a parameter of the distribution of the observed results of the Monte Carlo experiment. These results are subject to random error and therefore have a distribution. In the most frequent case, the parameter representing the desired result is

the mean of this distribution, that is to say, it is the average of the results that would be obtained by hypothetical indefinite repetitions of the Monte Carlo experiment. It is worth stating this situation explicitly, as follows. The desired answer is an unknown number θ, called the estimand. The Monte Carlo method consists of finding a function $t = t(\xi_1, \xi_2, \cdots)$ of a set of observed random numbers $\xi_1, \xi_2 \cdots$. When t is regarded as a function $t = t(*, *, \cdots)$, it is called an estimator of θ. The numerical value of t, obtained by evaluating the function for the numerical values of the observed $\xi_1, \xi_2 \cdots$, is called an estimate of θ. Replication of the experiment would involve different sets of numerical values for the $\xi_1, \xi_2 \cdots$; hence various experiments, for a fixed estimator t, would yield various estimates t. If the mean value of these estimates, written ξt, equals θ, the estimator is unbiased. The standard deviation of these estimates is called the standard error of the estimator and it measures the amount by which individual estimates may be expected to differ from the estimate. It is usually quite easy to discover unbiased estimators in a typical Monte Carlo problem. What requires greater skill, however, is to discover, among the unbiased estimators, an estimator with small, or even minimum, standard error.

Two general remarks follow from the above specification. First, ξt is the weighted average of estimate t taken over the sample space of observations $\xi_1, \xi_2 \cdots$, the respective weights being the probabilities of observing the respective $\xi_1, \xi_2 \cdots$. Hence t is an integral, usually over a space of many dimensions. Thus, the Monte Carlo method is essentially concerned with evaluating multidimensional integrals. This is not to say that it can be used only in problems of integration, but it does say that with any given problem, which may not at first sight be related to integration; one may expect to associate an underlying integration. Second, a small standard error will

occur if an estimator function t, which is relatively insensitive to variations in its arguments $\xi_1, \xi_2 \cdots$, can be chosen. For all estimates t will then be more or less equal. This is the idea that underlies the following technique, known as importance sampling. Suppose we have an unbiased estimator $t_1(\xi_1, \xi_2 \cdots)$ and $p_1(\xi_1, \xi_2 \cdots)d\xi_1 d\xi_2 \cdots$ is the probability density of the observed $\xi_1, \xi_2 \cdots$. Then, as indicated above,

$$\theta = \mathcal{E}t_1 = \int t_1(\xi_1, \xi_2, \cdots)\, p(\xi_1, \xi_2, \cdots)d\xi_1 d\xi_2 \cdots \tag{H-1}$$

define another estimator

$$t_2(\xi_1, \xi_2, \cdots) = \frac{t_1(\xi_1, \xi_2, \cdots)}{q(\xi_1, \xi_2, \cdots)} \tag{H-2}$$

where $q(\xi_1, \xi_2, \cdots)$ is a nonnegative function such that $p(\xi_1, \xi_2, \cdots) \bullet q(\xi_1, \xi_2, \cdots)$ is a probability density function, that is,

$$\int p(\xi_1, \xi_2, \cdots) \bullet q(\xi_1, \xi_2, \cdots)d\xi_1 d\xi_2 \cdots = 1 \tag{H-3}$$

From Equations (H-1), (H-2) and (H-3), θ can then be written as

$$\theta = \mathcal{E}t_1 = \int t_1(\xi_1, \xi_2, \cdots)\, p(\xi_1, \xi_2, \cdots)q(\xi_1, \xi_2, \cdots)d\xi_1 d\xi_2 \cdots \tag{H-4}$$

and therefore t_2 is also an unbiased estimator of θ, provided a way of sampling the $\xi_1, \xi_2 \cdots$ according to the new distribution $p(\xi_1, \xi_2, \cdots) \bullet q(\xi_1, \xi_2, \cdots)$. If q is chosen so that q is nearly proportional to t_1, then t_2 will not depend strongly upon $\xi_1, \xi_2, \cdots$, and therefore t_2 will have a small standard error. In order to be able to choose a nonnegative q nearly proportional to t_1, t_1 must be nonnegative for the most part; this can often be achieved by adding, if necessary, a known positive constant to t_1 and noting that $\mathcal{E}(t_1 + c) - c = \mathcal{E}t_1$. One may ask why the standard error cannot be reduced

to zero by taking q exactly proportional to t_1. The answer is that, in order to satisfy Equation (H-3), the following equation must be taken.

$$q(\xi_1, \xi_2, \cdots)=t_1(\xi_1, \xi_2, \cdots)/\theta \tag{H-5}$$

And the value of the constant of proportionality $1/\theta$ is actually unknown because the whole exercise is to determine the value of θ. In practice, it must be compromised between a q, which is rather like t_1, and a q, which will satisfy Equation (H-3). However, it is obvious that when t_1 is large, q and hence $p \bullet q$ ought to be large. That is to say, in using the new estimator t_2 an enhanced probability of sampling the regions, where t_1 is large, is needed. Large values of t_1 are made important, as it were, and hence the name importance sampling.

The foregoing discussion illustrates what seems at first sight to be a paradox in Monte Carlo work, namely, that having introduced randomness in the form of the observations, the effects of this randomness by a suitable choice of estimator is trying to be removed. However paradoxical this may seem, it is a fundamental precept of Monte Carlo work to regard randomness as a nuisance, to be avoided, suppressed, or reduced as much as possible.

In all cases, the Monte Carlo method is to be judged by its results. Thus, when a computer is used, it is convenient to replace the random numbers $\xi_1, \xi_2, \cdots$ by a set of so-called pseudo-random numbers. These are numbers that, although strictly deterministic in origin, pass for random numbers when subjected to certain conventional statistical tests of randomness. Much philosophical discussion has centred on the validity, and even the ethics, of forging random numbers in this way; such discussion is irrelevant if pseudo-random numbers yield as good final results as

real random numbers, and this is so often the case that little anxiety need be felt. A few kinds of random number generator can be found in Appendix C.

2 Interpolation of Functions of Several Variables

Suppose that a function $f(x_1, x_2, \cdots, x_n)$ of a large number of variables, say n=1000, have to be interpolated. To be specific and to keep the argument simple, imagine that f is a multi-linear function:

$$f(x_1, x_2, \cdots, x_n) = a + \sum_i a_i x_i + \sum_{i \neq j} a_{ij} x_i x_j + \cdots + a_{12\cdots n} x_1 x_2 \cdots x_n \qquad \text{(H-6)}$$

Here the coefficients ***a*** are unknown, the algorithm to determine it is to suppose that f can be calculated when all the x's are either 0 or 1. We have to determine the value of $f(p_1, p_2, \cdots, p_n)$ has to be determined, where the p_i are prescribed numbers between 0 and 1. In principle, the problem is elementary: $f(p_1, p_2, \cdots, p_n)$ is simply a linear combination, with known coefficients of the form $p_1 p_2 (1-p_3) p_4 \cdots (1-p_n)$, of the computable values of f at the vertices of the unit hyper-cube. In practice, however, this method is unworkable, because there are 2^n vertices and, when n=1000, the sum of 2^{1000} terms has to be found, each a product of 1001 quantities. Being confronted with much more data than what can be analysed at this stage, a method that uses only a tiny fraction of these data should be built first. In the absence of information that permits a rational method of selecting this fraction, there seems no alternative to random selection. Nevertheless, the prospects are not as gloomy as they may appear.

By Eq(H-6),

$$f(p_1, p_2, \cdots, p_n) = \sum r_1 r_2 \cdots r_n f(\delta_1, \delta_2, \cdots, \delta_n) \qquad \text{(H-7)}$$

where the sum is taken over all combinations of $\delta_1, \delta_2, \cdots, \delta_n = 0$ *or* 1 and $r_i = 1 - p_n$ *or* p_i according as $\delta_i = 0$ or 1. Now choose a set of numbers

$$\xi_i = \begin{cases} 0 & \text{with probability } 1 - p_i \\ 1 & \text{with probability } p_i \end{cases} \tag{H-8}$$

independently for $i = 1, 2, \ldots, n$. Thus, for example, if $p_i = 0.1$, ξ_i has a 90 per cent chance of being 0 and a 10 per cent chance of being 1. Then the probability

$$P[\xi_1 = \delta_1, \xi_2 = \delta_2, \cdots, \xi_n = \delta_n] = r_1 r_2 \cdots r_n \tag{H-9}$$

and hence, from Eq(H-7),

$$\xi f(\xi_1, \xi_2, \cdots, \xi_n) = f(p_1, p_2, \cdots, p_n) \tag{H-10}$$

Thus $f(\xi_1, \xi_2, \cdots, \xi_n)$ is an unbiased estimator of the estimand $f(p_1, p_2, \cdots, p_n)$. If replicating the procedure N times, the mean of the resulting N values of $f(\xi_1, \xi_2, \cdots, \xi_n)$ will be an unbiased estimator with standard error

$$\sigma = \left[\frac{1}{N} \text{var } f(\xi_1, \xi_2, \cdots, \xi_n \right]^{1/2} \tag{H-11}$$

Is σ small enough for this to be a useful procedure for moderate values of N? The answer depends, naturally enough, upon how well-behaved a function f is. We may describe the regularity of f in terms of the quantities $\Delta_i = \Delta_i(V)$, $(i = 1, 2, \cdots, n)$, where $\Delta_i(V)$ is the difference between the value of f at the vertex V of the unit hyper-cube and its value at the i^{th} of the n vertices adjacent to V. Let $M(r)$ be the smallest function of r such that

$$\sum_{(r)} \Delta_i^2 \le M(r) \tag{H-12}$$

holds at each vertex V of the hyper-cube, where $\sum_{(r)}$ denotes summation over any subset of r of the n possible values of i. Then,

$$\sigma \le N^{-1/2}\left[\frac{1}{4}M(1)+\frac{1}{2}\sum_{r=2}^{n}M(r)/r\right]^{1/2} \tag{H-13}$$

Two special but important cases of inequality (H-13) should be considered. The first case arises when the partial derivatives of f are all bounded. In this case, the differences Δ_i are all bounded, and let's describe $M(r)=Mr$, where M is a constant. Then inequality (H-13) yields

$$\sigma < \sqrt{(Mn/2N)} \tag{H-14}$$

and a satisfactorily small standard error can then be produced by taking N to be a suitably large multiple of n. The procedure can yield useful results if the number of dimensions is not too large, for example, only a few hundred. The second special case arises when $M(r)$ is independent of r, so that for each vertex we have

$$\sum_{(r)}\Delta_i^2 \le \sum_{i=1}^{n}\Delta_i^2 \le M \tag{H-15}$$

where M is once again a constant. Eq(H-15) is quite naturally satisfied by a reasonably regular function of several variables; it is a finite difference analogue of the supposition that grad f is uniformly bounded in magnitude. Inserting inequality (H-15) into (H-13), it can be deduced

$$\sigma < \sqrt{\left[(\frac{1}{2}+\log n)M/2N\right]} \tag{H-16}$$

The right-hand side of (H-16) is a very slowly increasing function of n. Even if n is prodigiously large, for example, if $n=10^{26}$, it is expected to get a satisfactorily small

standard error with a practicable number of replications, provided, of course, that one of these replications can actually be performed. This proviso is necessary because, in a replication, f must be evaluated at a vertex chosen at random; when n is prodigious, it is no longer a trivial matter to specify which vertex has been chosen. In fact, to specify a vertex amounts to specifying an integer having n binary digits, and the random procedure for choosing a vertex calls for the use of n random numbers. Thus, in the absence of special circumstances, the method is unworkable. Nevertheless, such special circumstances may arise as, for instance, when f is a symmetric function of its variables. In this case, only how many, rather than which, of the co-ordinates of the vertex are zero need to be specified. This can be done easily by choosing a random integer with a suitable binomial distribution between 1 and n. This is the sort of situation that can arise when dealing with functions of about 10^{26} variables in certain group-theoretic branches of atomic physics. A similar situation occurs when f is anti-symmetric, that is to say, when permutation of the variables can at most change the sign of f.

Thus far having considered only linear interpolation, what are the prospects for interpolation of higher order? In principle, polynomial interpolation is straightforward, for the multi-Lagrangian formula

$$f(x_1, x_2, \cdots, x_n) = \sum_j f(x_{1j}, x_{2j}, \cdots, x_{nj}) \prod_{i,k}{}' \frac{(x_i - x_{ik})}{(x_{ij} - x_{ik})} \qquad \text{(H-17a)}$$

where j ranges over all data points, and Π' denotes the product omitting terms of the type $(x_i - x_{ik})/(x_{ij} - x_{ik})$. Eq(H-17a) may be more concisely written as

$$f = \sum_j L_j f_j = \sum_j p_j (L_j f_j / p_j) \qquad \text{(H-17b)}$$

Thus, if choosing the *j*th data point with probability p_j, $L_j f_j / p_j$ can be taken as an unbiased estimate of the interpolate f. Here the nonnegative p_j are quite arbitrary, subject only to

$$\sum_j p_j = 1 \tag{H-18}$$

so that the p_j need to be selected in a manner that will both reduce the standard error according to the principles of importance sampling, and also make $L_j f_j / p_j$ a simple quantity to calculate.

In practice, however, it does not seem easy to meet these desiderations. Although the identity $\sum_j L_j = 1$ holds in tempting conformity with Eq(H-18), unlike what could be done in the linear case, $p_j = L_j$ can no longer be taken because there will always be some negative L_j. Not only does change of sign in the L_j prevent this simplicity, but it also tends to inflate the standard error. For,

$$\begin{aligned} \text{var}\, f &= \sum_j p_j (L_j f_j / p_j)^2 - (\sum_j L_j f_j)^2 \\ &= \left[\sum_j \left(\frac{|L_j f_j|}{\sqrt{p_j}} \right)^2 \right] \left[\sum_j (\sqrt{p_j})^2 \right] - (\sum_j L_j f_j)^2 \\ &\geq (\sum_j |L_j f_j|)^2 - (\sum_j L_j f_j)^2 \end{aligned} \tag{H-19}$$

By Cauchy's inequality, the right-hand side of inequality (H-19) thus presents a minimum variance, which cannot be improved upon however ingeniously of sampling. This minimum is attained if and only if it is sampled with probabilities

$$p_j = \frac{|L_j f_j|}{\sum_j |L_j f_j|} \tag{H-20}$$

Now it may easily happen that the right-hand side of inequality (H-19) is very large in comparison with f. Indeed, this will be true if all the f_j are roughly of the same size and if $\sum_j |L_j|$ is very much larger than 1. For simplicity, it can be supposed that the data points consist of all possible combinations of the form $(x_1', x_2', \cdots, x_n')$, where x_i' belongs to a given set of points Λ_i. Writing $L_j^{(i)}$ for the univariate Lagrangian coefficients determined by Λ_i, then,

$$\sum_j |L_j| = \prod_{i=1}^{n} \left\{ \sum_j |L_j^{(i)}| \right\} \tag{H-21}$$

If Λ_i consists of more than two points, $\sum_j |L_j^{(i)}| > 1$. Consequently, the left-hand side of Eq(H-21) will, in general, increase exponentially with *n*, and will be very much larger than 1 unless *n* is really small, for example, $n < 10$.

It can be concluded that a straightforward Monte Carlo exploitation of the non-linear multi-Lagrangian interpolation formula will be wholly unsatisfactory unless the number of variables for which non-linear interpolation is essential is very small, for example, ten at most. Moreover, when *n* is as small as this, a direct evaluation of the classical Eq(H-17a) is quite feasible. On the other hand, linear interpolation by Monte Carlo methods seems to be a practicable and attractive approach for functions of many variables, and possesses the appealing feature that any interpolation coefficients does not have to be calculated since this is automatically dealt with, without calculation, by the sampling procedure.

This difference between linear and nonlinear interpolation, so far as Monte Carlo methods are concerned, may seem surprising when we recall Aitken's well-known technique of effecting a nonlinear interpolation by a succession of linear

interpolations and extrapolations (Hartree, D. R., 1958). Why can not exploit this device? The answer is that, although the Monte Carlo method will handle linear interpolations, it will fail on linear extrapolations since for these negative L_j are required, and the difficulties inherent in inequality (H-19) and Eq(H-21) will present themselves. Is there no alternative? Suppose we abandon Lagrangian interpolation and replace Eq(H-17b) by some general linear form

$$f = \sum_j w_j f_j \tag{H-22}$$

The character of the interpolation problem seems to dictate a linear structure such as Eq(H-22). This will lead to the difficulties of relations of Eq(H-19) and (H-21) if some of the w_j are negative. On the other hand, some w_j must be negative unless

$$\min_j f_j \le f \le \max_j f_j \tag{H-23}$$

Since the original function of f may easily violate Eq(H-23), the only hope seems to be to devise a preliminary transformation of the function f such that the transformed function satisfies Eq(H-23). This being done, an interpolation formula (as Eq(H-22)) can be introduced, in which the w_j are positive and satisfy

$$\sum_j w_j = 1 \tag{H-24}$$

Further research seems to be required to see whether this scheme is feasible and whether it leads to a Monte Carlo result with a suitably small standard error. At present, the problem of non-linear interpolation in a function of a few hundred variables appears to be unsolved. The complete set of data for a non-linear interpolation consisting of all possible combinations of the form $(x_1', x_2', \cdots, x_n')$

where x_i' belongs to a given set Λ_i, its analysis to give slopes and curvatures in various co-ordinate directions, and the incorporation of these slopes and curvatures into the interpolation formula yielding the required interpolate is essentially what a statistician would call a complete factorial analysis. The complete formula takes account of all interactions of all factors, whereas in practice many of the high-order interactions are unimportant. When a statistician encounters this situation he curtails his data and confounds the high-order with the low-order interactions (D.B.Lewis, et al., 1995). Can a corresponding advantage be gained in non-linear interpolation problems by confounding? Some problems that require attention in the case of linear interpolation are the following. Is Eq(H-12) a satisfactory criterion, and how does one estimate the function M(r) for a given f? To what extent can importance sampling or other variance-reducing techniques be used? In particular, how does one determine the importance function, and to what extent can one usefully modify it sequentially as the sampling provides information about the nature of f? What sort of special features in f will allow specification of vertices when n is very large?

3 Calculations with Large Matrices

The following speculative material serves as an introduction to the discussion in the next section on methods for reducing the number of variables. Calculations with large matrices create difficulties for computers, not only because of the time required to carry through the relevant arithmetic, but also because considerable storage space is required. For instance, a square matrix of order 10^5 requires a store of 10^{10} words to hold it. Could the calculations be effected to a reasonable degree of accuracy using only a sample of the elements of such matrices?

Suppose it is wished to calculate the element c_{ij} of a matrix product C=AB, where $A=(a_{ij})$ and $B=(b_{ij})$ are matrices of order $l\times m$ and $m\times n$. It is known that

$$c_{ij}=\sum_{k=1}^{m}a_{ik}b_{ki} \tag{H-25}$$

Hence, if chosen uniformly at random N different elements from the *i*th row of A, say, $a_{ik_1},\cdots,a_{ik_N}$, together with the corresponding elements of the *j*th column of B,

$$r_{ij}=\frac{m}{N}\sum_{r=1}^{N}a_{ik_r}b_{k_r j} \tag{H-26}$$

will be an unbiased estimator of c_{ij}. The standard error of r_{ij} is

$$\left[\frac{(m-N)}{N(m-1)}\right]^{1/2}\left[m\sum_{k=1}^{m}a_{ik}^2b_{kj}^2-\left(\sum_{k=1}^{m}a_{ik}b_{kj}\right)^2\right]^{1/2} \tag{H-27}$$

The second bracketed factor will be of the same order of magnitude as c_{ij} itself, if the products $a_{ik}b_{kj}$ are of roughly equal magnitudes. For instance, if $0<\lambda\le a_{ik}b_{kj}\le\Lambda$ for $k=1,2,\cdots,m$, then

$$\left[m\sum_{k=1}^{m}a_{ik}^2b_{kl}^2-\left(\sum_{k=1}^{m}a_{ik}b_{kl}\right)^2\right]^{1/2}\le c_{ij}(\Lambda-\lambda)/(2(\Lambda\lambda)^{1/2}) \tag{H-28}$$

Eq(H-26) may thus be expected to have an acceptable standard error (for example, when N=1000), and the Monte Carlo method will be worth considering if $m\ge 10^5$, for example, since it will reduce the computational labour at least one hundred fold without unreasonable loss of accuracy.

Furthermore, if m is extremely large, the statistical errors in the Monte Carlo result actually may be less than the accumulated rounding errors in an "exact" calculation of the full Eq(H-25).

Then, consider the question of finding a dominant eigenvector of a large matrix A, that is, the eigenvector corresponding to the eigenvalue of greatest modulus. If u_n is an arbitrary vector and $u_{n+1} = Au_n$, it is obvious that, as $n->\infty$, u_n becomes asymptotically proportional to the dominant eigenvector. It may therefore be desired exploit the foregoing technique for matrix multiplication in order to estimate u_n. Consider the following procedure. Suppose an estimate of a scalar multiple of u_n has been introduced, for example,

$$V_n = \{v_{n1}, v_{n2}, \cdots, v_{nm}\} \qquad \left(\sum_{i=1}^{m} v_{ni} = 1\right) \tag{H-29}$$

Take $v_{n+1,i} = \alpha a_{ik} v_{nk}$ where k is a randomly chosen integer between 1 and m, and α is chosen so that $\sum_{i=1}^{m} v_{n+1,i} = 1$, then $w = \frac{1}{N}\sum_{n=1}^{N} V_n$ will be a slightly biased estimator of the dominant eigenvector. The bias is due to normalising the vector by multiplying by α and will be of the order of $1/m$, which will be negligible for a large matrix. One can modify the procedure so that the bias is removed (see below), but such modifications are probably not worthwhile.

The process will fail, of course, if it is unable at any stage to determine α because the sum of the elements of v_n is zero. Such an accident, with probability one, will occur sooner or later if there is a zero element somewhere in A, for example; it cannot occur, however, if all the elements of A are positive and if $u_0 = \{1,1,\cdots,1\}$ has been taken. There is, indeed, no loss of generality in starting with this u_0 since, in any case,

u_0 may be arbitrary. Furthermore, even if starting with some other u_0, at some point there will be a probability at one stage where k=1 for all i, whereupon v_n will be proportional to the first column of A and all subsequent v_n will be independent of u_0. Since, indeed, the event of k=1 for all i has a probability m^{-m} at each stage, the whole process is a renewal process. Therefore, the standard error of $w = \frac{1}{N}\sum_{n=1}^{N} V_n$ will be asymptotically proportional to $N^{1/2}$ for large N under weak general conditions, for example, when all the elements of A are positive and all the elements of v_n are consequently bounded. How large N must be before this asymptotic proportionality becomes effective is a rather difficult question which deserves further study *j* it seem quite plausible, though, that N need not be nearly as large as m if the elements of A are not too dissimilar. As stated above, one can elaborate the procedure and make it unbiased. One way of doing this is to carry out two separate and independent experiments, and to use the α's found in each experiment to normalise the vectors of the other experiment. The two experiments will have to be carried through simultaneously, which doubles the amount of storage required in the computer and, since the bias of a single experiment is small, the additional complications of a double experiment seem scarcely profitable.

4 Reduction of Multivariable Functions

A multivariable function is handled more easily when it can be expressed to an adequate degree of approximation in the form

$$f(x_1, x_2, \cdots, x_k) = \sum_{r=1}^{R} g_{r1}(x_1) g_{r2}(x_2) \cdots g_{rk}(x_k) \qquad \text{(H-30a)}$$

C. D. Allen (1959) has given a method of achieving such a representation in the case k=2, with an indication of possible extensions to larger values of k, following him, it is easily to begin with the case k=2, which can be written in the form

$$f(x,y)=\sum_{r=1}^{R} g_r(x)h_r(y) \qquad \text{(H-30b)}$$

Here it needs to be found, with as small a value of R as possible, functions g_r and h_r such that Eq(H-30b) holds approximately at all points of the rectangular array (x_i, y_i) ($i=1,2,\cdots,m$ and $j=1,2,\cdots,n$, where $f_{ij}=f(x_i,y_j)$ is given. Thus, the $m\times n$ matrix $F=(f_{ij})$ is given. For each value of r column vectors have to be determined,

$$\begin{aligned} g_r &= \{g_r(x_1), g_r(x_2), \cdots, g_r(x_m)\} \\ h_r &= \{h_r(y_1), h_r(y_2), \cdots, h_r(y_n)\} \end{aligned} \qquad \text{(H-31)}$$

Allen's procedure is equivalent to the following recipe. Take g_r to be the column eigenvector of unit length corresponding to the rth largest eigenvalue of the $m\times m$ matrix FF', where the prime denotes transposition; then take $h_r=F'g_r$. This process may be carried out as follows. When g_1, g_2, ... , g_{r-1} and $h_1, h_2, \cdots, h_{r-1}$ have been determined, $F_r=F-\sum_{s=1}^{r-1} g_s h_s'$ can then be calculated, and then find g_r as the dominant eigenvector of F_rF_r', and $h_r=F_r'g_r$. The process will be stopped when F_r is negligibly small, and this gives R.

Thus, it is enough to consider a method of estimating a dominant eigenvector of a matrix FF', where F is given. This can be done by combining the two techniques of the previous section. Starting from $V_0=\{1,1,\cdots,1\}$, calculate a sequence of vectors V_v, where

$$v_{v+1,j} = \alpha f_{ik} f_{lk} v_{vl} \quad (i = 1,2,\cdots,m) \tag{H-32}$$

and k and l are integers chosen uniformly at random in the ranges 1 to n and 1 to m, respectively, and α is a normalising constant such that $\sum_{i=1}^{m} v_{v+1,j} = 1$. For a suitable large value of N, the dominant eigenvector of FF' can be estimated as

$$g = \frac{\beta}{N} \sum_{v=1}^{N} V_v \tag{H-33}$$

where β is such that g has unit length. Having done this, the product $h = F'g$ can then be estimated by the method of matrix multiplication discussed in the previous section.

Next, the multidimensional case Eq(H-30a) with an arbitrary value of k need to be considered. The data will consist of a tensor

$$f_{i_1 i_2 \cdots i_k} = f(x_{1i_1}, x_{2i_2}, \cdots, x_{ki_k}) \quad (i_s = 1,2,\cdots,m_s;\ s = 1,2,\cdots,k) \tag{H-34}$$

Representing the values of the function at a total of $m_1 m_2 \cdots m_k$ points of a rectangular array. By treating the x and y of Eq(H-30a) as portmanteau variables, it can be seen that, if

$$g_{1s}(x_{i_s}) = g_{i_s}^{(s)} \qquad (i_s = 1,2,\cdots,m_s) \tag{H-35}$$

then

$$g_{i_s}^{(s)} = \lim_{N \to \infty} \frac{\beta}{N} \sum_{v=1}^{N} v_{v,i_s}^{(s)} \tag{H-36a}$$

$$g_{i_k}^{(k)} = g_{i_1}^{(1)} g_{i_2}^{(2)} \cdots g_{i_{k-1}}^{(k-1)} f_{i_1 i_2 \cdots i_k} \tag{H-36b}$$

where

$$v_{v+1,i_1}^{(1)} = \alpha f_{i_1 i_2 \cdots i_k} f_{j_1 i_2 \cdots i_k} v_{v,j_1}^{(1)} \tag{H-37a}$$

$$v_{v+1,i_s}^{(s)} = \alpha g_{i_1}^{(1)} g_{i_2}^{(2)} \cdots g_{i_{s-1}}^{(s-1)} f_{i_1 i_2 \cdots i_k} g_{j_1}^{(1)} g_{j_2}^{(2)} \cdots g_{j_{s-1}}^{(s-1)} f_{j_1 j_2 \cdots j_s i_{s+1} \cdots i_k} v_{v,j_s}^{(s)} \quad (s = 2,3,\cdots,k-1) \tag{H-37b}$$

In Eq(H-36b) and Eq(H-51a, b), the summation convention operates over all repeated suffixes; the α's are normalising constants such that

$$\sum_{i_s=1}^{m_s} v_{v+1,i_s}^{(s)} = 1 \qquad (s = 1,2,\cdots,k-1) \tag{H-38}$$

and the β's are normalising constants such that

$$\sum_{i_s=1}^{m_s} (g_{i_s}^{(s)})^2 = 1 \qquad (s = 1,2,\cdots,k-1) \tag{H-39}$$

As set out here, these relations are exact. Their evaluation in this exact form, however, will be impracticable as a rule because of the large number of terms implied by the summation convention. Hence for computing purposes we replace the summation convention by sampling, and interpret the right-hand side of Eq(H-37a) as a single term in which the suffixes $j_1, i_2, i_3, \cdots, i_k$ are integers chosen uniformly at random from the respective ranges 1 to m_s (s = 1, 2, , ..., k). Similarly, for fixed s in Eq(H-37b), $i_1, i_2, \cdots, i_{s-1} i_{s+1}, \cdots, i_k$ and $j_1, j_2, \cdots, j_s$ are integers chosen uniformly at random from the appropriate ranges. In Eq(H-36a), remove the operator $\lim_{N\to\infty}$ and take N finite but sufficiently large, and in Eq(H-36b) sample a sufficient number of random combinations of $i_1, i_2, \cdots, i_{k-1}$.

To get the corresponding expressions for the functions $g_{2s}(x)$, take $f_{i_1 i_2 \cdots i_k} - g_{i_1}^{(1)} g_{i_2}^{(2)} \cdots g_{i_k}^{(k)}$ in place of $f_{i_1 i_2 \cdots i_k}$ and so on for $g_{rs}(x)$ with $r > 2$. An

algebraic approach would require a manageable value of $\prod_{s=1}^{k} m_s$, whereas the Monte Carlo approach requires only $\sum_{s=1}^{k} m_s$ to be manageable, and therefore appears preferable except when k is very small.

5 Evaluation of Multidimensional Integrals

If R is a region of multidimensional space in which x is a typical vector variable, the integral

$$I = \int_R f(x)dx \tag{H-40}$$

has for an unbiased estimator

$$\sum_j w_j f(\xi_j) \tag{H-41}$$

where ξ_j are points chosen uniformly at random in R, and the w_j are weights with unit sum. The standard error of estimator (H-41) can be reduced by suitable importance sampling for the ξ_j if there is adequate information about f. A different kind of procedure, to be used with or without importance sampling, exploits the fact that the ξ_j in Eq(H-41) need not be independent. In fact, it is reasonable to choose the ξ_j in sets such that large values of $f(\xi_j)$ offset small values, thus reducing the effects of random variation. A fuller account of this technique may be found elsewhere. (Neal J. Adams, 1985, T. S. Motzkin, 1956, D. R. Hartree, 1958)

One possibility, which does not appear to have received adequate attention as yet, is to replace I by a numerical integration formula (the integrated form of a Lagrangian

interpolation formula) and then sample various terms of this formula in the manner already described for interpolation by Monte Carlo methods. it has been shown that there are difficulties in non-linear interpolation due to negative Lagrangian coefficients. However, the integrated versions do not necessarily suffer from this difficulty, for example, there are no negative coefficients in the *n*-point Lagrangian integration formula for $n \leq 8$, and also for $n = 10$. Here n is the number of points in each dimension, so that if k is the dimensionality of the space in Eq(H-40), it can be sampled from a set of n^k points with probabilities proportional to the products of the respective one-dimensional *n*-point Lagrangian integration coefficients (A. N, Lowan, 1944).

6 Solution of Multidimensional Integral Equations and Partial Differential Equations

The following standard Monte Carlo procedures are included for reference purposes.

Suppose the following equation need to be solved

$$M(x) = M_0(x) + \int_R K(x,y)M(y)dy \qquad (x \in R) \tag{H-42}$$

where x and y are vectors, M_0 and K are known functions, R is a given region of space, and M is to be determined. Let $P(x,y)$ be any function which satisfies

$$P(x,y) > 0 \quad \text{and} \quad \int \mathrm{P(x,y)dy} < 1 \tag{H-43}$$

R

Then define

$$\begin{aligned} P(x) &= 1 - \int_R P(x,y)dy \\ W(x) &= M_0(x) / P(x) \\ W(x,y) &= K(x,y) / P(x,y) \end{aligned} \tag{H-44}$$

The integral Eq(H-41) becomes

$$M(x) = P(x)W(x) + \int_R P(x,y)W(x,y)M(y)dy \tag{H-45}$$

Consider a random walk that starts at a point $x_0 = x$ of R, and has successive steps $x_0 \to x_1 \to \cdots$ governed by the following law: when the walk is at x_n, it has a probability $P(x_n, x_{n+1})dx$ of taking its next step to the element of volume dx at x_{n+1} in R and a probability $P(x_n)$ of leaving R. If the walk leaves R after being at a point x_N of R, a score can be associated with this random walk

$$S(x) = W(x_0, x_1)W(x_1, x_2)\cdots W(x_{N-1}, x_N)W(x_N) \tag{H-46}$$

It is easy to see from Eq(H-45) that S(x) is an unbiased estimator of the required function M(x). The question now arises of how to choose the arbitrary function P(x, Y) so that the estimator S(x) has a small standard error. It can be shown quite easily that if $P(x,y)$ is proportional to $K(x,y)M(y)/M(x)$, then S(x) has a zero standard error. Of course, one cannot satisfy this criterion in practice without knowing the desired solution $M(x)$; furthermore, the ratio $K(x,y)M(y)/M(x)$ may not be of constant sign, so that the first of the two inequalities (H-43) would prevent strict proportionality. Nevertheless, if there is a rough idea of the solution $M(x)$, a $P(x,y)$ can often be chosen which will be approximately proportional to the ratio $K(x,y)M(y)/M(x)$, and such a choice should give a small standard error for S(x).

Turning to partial differential equations, let us consider for the sake of simplicity the multidimensional Laplace equation

$$\nabla^2 \phi = 0 \tag{H-47}$$

Other elliptic partial differential equations can be handled by straightforward modification of the following recipe. Imagine that Eq(H-47) has to be solved in a region R with given values of ϕ on the boundary of R. Let $\sum_x$ denote a sphere within R centred at a point x, and let $P(x,y)$ denote a probability density function over the points on the surface of $\sum$. Starting from $x_0 = x$ perform a random walk $x_0 \to x_1 \to \cdots$, until the walk reaches the boundary of R at, say, x_N where, given x_n, the next step is taken to a point x_{n+1} on the surface of $\sum_{x_n}$ chosen according to the density $P_n(x_n, x_{n+1})$. Associate with this walk the score

$$T(x) = \frac{V_0(x_0)}{P_0(x_0,x_1)} \cdot \frac{V_1(x_1)}{P_1(x_1,x_2)} \cdots \frac{V_{N-1}(x_{N-1})}{P_{N-1}(x_{N-1},x_N)} \cdot \phi(x_N) \qquad \text{(H-48)}$$

where $V_i(x_i)$ is the reciprocal of the surface area of $\sum_{x_n}$, and $\phi(x_N)$ is the given boundary value. Then, because ϕ is harmonic, T(x) is an unbiased estimator of $\phi(x)$. To get a small standard error on the basis of rough knowledge of the behaviour of ϕ, choose the $P(x,y)$ as nearly as possible proportional to $\phi(y)$. Fuller information about the foregoing matters has been presented by Curtiss (J. M. Hammersley, 1956).

7 Minimal Irregularities of Distribution

In carrying out a Monte Carlo experiment on a digital computer, one normally uses pseudo-random or quasi-random numbers in place of random numbers. O. Taussky and J. Todd (1956) and R.D. Richtmyer (1958) have discussed the generation and properties of these numbers. One of the most popular sequences of pseudo-random numbers is

$$\xi_n = \{au^n / v\} \tag{H-49}$$

where a, u, and v are integers, and {x} denotes the fractional part of x. A well-known example of a sequence of quasi-random numbers is $\xi_n = \{\alpha n\}$, where α is irrational. Such numbers are easy to generate on a computer. Pseudo-random numbers often satisfy a battery of the standard statistical tests for randomness, even though they are strictly deterministic; this fact has been advanced as a justification for their use. Quasi-random numbers do not usually pass these tests quite so easily, but both quasi-random as well as pseudo-random numbers seem to provide the right final answers when used in a Monte Carlo procedure. This latter criterion is much more important than the satisfaction of statistical tests, and it seems quite conceivable that quasi-random numbers may be more satisfactory than genuinely random ones. This would be true if, while still yielding unbiased answers, they afforded smaller standard errors than random numbers do. As a rule, the standard error, obtained by employing N random numbers, behaves like $N^{-1/2}$. On the other hand, Richtmyer has shown that, in theory, the corresponding standard error arising from N quasi-random numbers behaves asymptotically like N^{-1}. However, it has also been shown that, in practical examples, the performance of quasi-random numbers is not as markedly superior to random numbers as the theoretical asymptotic results would indicate. Presumably this is because the true asymptotic behaviour is not realised until $N \geq N_0$, for example, where N_0 is impracticably large for the quasi-random numbers studied. Nevertheless, it seems quite possible that N_0 may depend quite critically upon the type of quasi-random number used, for example, upon the value of a in $\xi_n = \{\alpha n\}$; hence there is a definite need for further research in this field.

Now, consider the evaluation of

$$I = \int_C f(x)dx \tag{H-50}$$

where C is a unit hyper-cube. It can be estimated by a formula $\sum w_i f(x_i)$ and, in the absence of special information about f, it can be done no more than take the weights w_i equal. Since f may be approximated by a sequence of simple integral functions, that is, functions which take only finitely many different values, and since the sets on which f is constant can be built up by addition and intersection of rectangles, it will be a reasonable criterion if we judge the efficiency of estimation of integral (H-50) merely by its efficiency when f is the indicator function of a rectangle in C. Thus, the points x_i needs to be chosen so that, as nearly as possible, the proportion that falls in any given sub-rectangle of C equals the volume of this rectangle. There is, quantitatively, no real loss of generality in assuming that one corner of this rectangle coincides with a corner of the hyper-cube and that its sides re parallel to the sides of the hyper-cube. Thus, if working with a sequence of N points in a unit hyper-cube of k dimensions, the following equation may be used as a criterion

$$J = \int_0^1 \cdots \int_0^1 \{S(x_1, x_2, \cdots, x_k) - Nx_1 x_2 \cdots x_k\}^2 dx_1 dx_2 \cdots dx_k \tag{H-51}$$

where $S(x_1, x_2, \cdots, x_k)$ is the number of these N points that fall in the rectangle whose upper right vertex is at $(x_1, x_2, \cdots, x_k)$.

Consider first how a classical numerical integration formula will rate against the Eq(H-51). If we take $N = m^k$, where m is an integer, and spread the integration points over the hyper-cubical lattice

$$x_i = (r_i + 1/2)/m \quad (r_i = 0,1,\cdots,m-1; i = 1,2,\cdots,k) \tag{H-52}$$

then,

$$J = \left[(m^2 + 1/2)/3\right]^k - 2\left[(m^2 + 1/8)/3\right]^k + \left[m^2/3\right]^k \qquad \text{(H-53)}$$
$$\sim (k/(3^k * 4)N^{2(k-1)/k} \qquad as \quad N$$

A detailed proof of Eq(H-53) is shown by J. M. Hammersley (1956).

A crude Monte Carlo method in which the N points are distributed uniformly at random over the unit hyper-cube will given an average value of J equal to

$$\varepsilon J = \left(\frac{1}{2^k} - \frac{1}{3^k}\right)N \qquad \text{(H-54)}$$

Thus, in one dimension, the classical formula $(J \sim 1/12)$ is much better than the crude Monte Carlo method ($\varepsilon J = \frac{1}{6} N$); in two dimensions, the classical formula, with $J \sim (1/18)N$, is $2\frac{1}{2}$ times as good as the crude Monte Carlo method for which $\varepsilon J = (5/36)N$; in three or more dimensions, though, the crude Monte Carlo method is much superior for large N. It should be emphasised that these qualitative results apply when the integrand is unknown and that, if having information about this integrand, a more efficient classical formula suited to this information is able to be selected. Furthermore, it may very justifiably suspect that there exists a deterministic way of choosing the N points in the hyper-cube that will be better than the Monte Carlo method. For instance, a set of N points that minimises J in Eq(H-52) should be an efficient set for the general body of integral functions of bounded variation. Unfortunately, it is not known what sets have this property. Indeed, it is not even known what is the greatest lower bound of J. The best result available in the literature is due to K.F. Roth (1954), who proved the existence of a positive constant c_k such that

$$J > c_k (\log N)^{k-1} \qquad \text{(H-55)}$$

For two dimensions, there is a result in the opposite direction due to J.C. Van der Corput (1935), who showed that there exists a sequence of N points for which

$$\sup_{\substack{0\le x_1\le 1\\ 0\le x_2\le 1}} \left| S(x_1,x_2) - N x_1 x_2 \right| \le A \log N \tag{H-56}$$

where A is a positive constant. The criterion involved in inequality (3-70) is a minimal one, stronger than the criterion provided by Eq(3-65). Since it gives an upper bound not too dissimilar to the lower bound given by Eq(3-69), Van der Corput's sequence should be a good one for evaluating two-dimensional integrals. The recipe for generating it is as follows. Write down the first N natural numbers in binary notation:

$$\begin{array}{c} 1 \\ 10 \\ 11 \\ 100 \\ 101 \\ \vdots \end{array} \tag{H-57}$$

Place a binary point after each such integer, and then read the digits backward:

$$\begin{array}{c} 0.1 \\ 0.01 \\ 0.11 \\ 0.001 \\ 0.101 \\ \vdots \end{array} \tag{H-58}$$

Then the nth point of Van der Corput's sequence has co-ordinates n/N, y_n, where y_n is the binary number occurring in the nth row of the sequence provided by Eq(H-58). Nothing is known of any corresponding sequence in more than two dimensions; however, the following conjectural generalisation is worth advancing. Let $\pi_2 = 2$,

$\pi_3 = 3$, $\pi_4 = 5$, $\cdots, \pi_k$ be the first $k-1$ prime numbers. In place of sequence provided by Eq(H-57), write down the first N natural numbers in π_i-nary notation $(2 \le i \le k)$ and then, in place of sequence provided by Eq(H-58), read these numbers backward as π_i-nary decimals, taking the nth number in such a list to be y_{ni}, where $(n/N, y_{n2}, y_{n3}, \cdots, y_{nk})$ is the nth point of the sequence in the k-dimensional unit hyper-cube. It would be interesting to know if, for this sequence,

$$\sup_{\substack{0 \le x_i \le 1 \\ i=1,2,\cdots k}} |S(x_1, x_2, \cdots, x_k) - N x_1 x_2 \cdots x_k| \tag{H-59}$$

does not exceed a power of $\log N$ as $N \to \infty$.

Roth[23] has Shown that there is a connection between the two-dimensional problem and a similar one-dimensional problem originally considered by van der Corput (1945). Let $\sum_1$ and $\sum_2$ denote sets of N points in the unit interval and the unit square, respectively. Suppose the points of $\sum_1$ are arranged in some fixed arbitrary order, and let $\sum_1^n$ denote the first n points $(1 \le n \le N)$. Let $S_n(x)$ denote the number of points in $\sum_1^n$ whose co-ordinates do not exceed x, for $0 \le x \le 1$. Then there exists a function A_N which is bounded and has a bounded reciprocal, such that

$$\inf_{\sum_1} \sup_{\substack{0 \le n \le N \\ 0 \le x \le 1}} |S_n(x) - nx| = A_N \inf_{\sum_2} \sup_{\substack{0 \le x \le 1 \\ 0 \le y \le 1}} |S(x,y) - Nxy| \tag{H-60}$$

where S(x, y) is the number of points of $\sum_2$ in the rectangle whose upper right comer is at (x, y). Van der Corput (1935) defined a just distribution of N points to be one for which the left-hand side of Eq(H-60) was bounded for all N, and it was conjectured that no just distribution existed. This conjecture was first proved by T.

van Aardenne-Ehrenfest (1949) who subsequently proved that the left-hand side of Eq(H-60) increases at least as fast as a multiple of $\log\log N / \log\log\log N$. Roth's results provided by Eq(H-55) and (H-60) improve the results further and show that the left-hand side of Eq(H-60) increases at least as fast as a multiple of $(\log N)^{1/2}$. The gap between $\sqrt{(\log N)}$ and $\log N$, the latter given by van der Corput's result Eq(H-56), remains unimplemented at present.

There is also another connection between one- and two-dimensional problems. Van der Corput proved that a necessary and sufficient condition for a sequence of points (α_n, β_n) (n=1, 2, ...) to be uniformly distributed over a square was that $\{u\alpha_n + v\beta_n\}$ should be uniformly distributed over the unit interval for every pair of integers u, v not both zero. Subsequently W.J. Coles (1957) gave the following quantitative expression to this result. Let $NF^{(N)}(x_0, x_1; y_0, y_1)$ denote the number of points in the set (α_1, β_1), (α_2, β_2), ..., (α_N, β_N) which fall in the rectangle whose opposite corners are (x_0, y_0) and (x_1, y_1). Define

$$D^{(N)} = \sup_{x_0, x_1, y_0, y_1} \left| F^{(N)}(x_0, x_1; y_0, y_1) - (x_1 - x_0)(y_1 - y_0) \right| \qquad \text{(H-61)}$$

Similarly write $NF_{uv}^{(N)}(x_0, x_1)$ for the number of points in the set

$$\{u\alpha_n + v\beta_n\} \qquad (n = 1, 2, \cdots, N)$$

which fall in the interval (x_0, x_1), and define

$$D_{uv}^{(N)} = \sup_{x_0, x_1} \left| F_{uv}^{(N)}(x_0, x_1) - (x_1 - x_0) \right| \qquad \text{(H-62)}$$

Then Coles proves the existence of an absolute constant C such that

$$D^{(N)} \le \varepsilon + C\left[D_{01}^{(N)} + D_{10}^{(N)} + \sum_{\substack{(u,v)=1 \\ u>0,\, v\neq 0}} D^{(N)} \min\left\{ \frac{1}{|uv|}, \frac{1}{|\varepsilon uv|^2} \right\} \right] \tag{H-63}$$

for any $\varepsilon > 0$, where $(u,v)=1$ denotes that u and v are coprime. The inequality (H-63) implies Van der Corput's necessary and sufficient condition,

Another line of inquiry on irregularities of distribution in one dimension concerns not the irregularities of the positions of the points of a sequence in a unit interval but the irregularities of the intervals between successive points, Suppose that we have an ordered infinite sequence $\sum$ of points in a unit interval, and that the first $n-1$ of these points divides the interval into n subintervals of which the shortest and the longest have lengths u_n and v_n respectively. For this situation, then, N.G. De Bruijn andP. Erdos (1949) proved the following inequalities:

$$\begin{aligned} &\limsup_{n\to\infty} nv_n \ge 1/\log 2 \\ &\liminf_{n\to\infty} nu_n \ge 1/\log 4 \\ &\limsup_{n\to\infty} v_n/u_n \ge 2 \end{aligned} \tag{H-64}$$

and it was shown by H.F, Trotter et al. (1956) that these results were the best possible, in that equality holds throughout the relations Eq(H-64) if the nth point of $\sum$ has co-ordinate

$$x_n = \{\log_2(2n+1)\} \qquad (n = 1,2,\cdots) \tag{H-65}$$

De Bruijn and Erdos also point to a connection between the work of van Aardenne-Ehrenfest (1949) and a generalisation of their own results, in which u_n and v_n are replaced by the shortest and longest total lengths of r adjacent intervals. From the statistical and Monte Carlo point of view it is interesting to note what can be said

about the variance V_n of the lengths of the n intervals created by the first $n-1$ points of $\sum$. Therefore, it can be proved that for any $\sum$,

$$\limsup_{n\to\infty} nV_n \geq \frac{1}{2(\log 2)^2} - 1 \tag{H-66}$$

and that this result is the best possible in the sense that

$$nv_n \leq \frac{1}{2(\log 2)^2} - 1 \qquad (n = 1,2,\cdots) \tag{H-67}$$

for the particular $\sum$ generated by Eq(H-65). Eq(H-65) is notable not merely for the ease with which it can be generated on a binary digital computer but also for its similarity to Eq(H-35). These three equations all have the form $\{f(n)\}$, where f is a simple function of n. Although there is much work in the literature on the coarser properties of sequences of the type $\{f(n)\}$, more research is needed on the finer properties.

Mathematics and Computers in Science and Engineering
A Series of Reference Books and Textbooks

Recent Advances in Applied and Theoretical Mathematics

Edited by Nikos E. Mastorakis
Military Institutions of University Education
Hellenic Naval Academy, Greece

World Scientific
Engineering Society **http://www.worldses.org**

Using Asymptotic Important Sampling to the Tail Model in Quantitative Risk Assessment

X.ZHANG, M.MIHSEIN, F.R.HALL, K.KIBBLE
School of Engineering and Built Environment,
University of Wolverhampton,
Wulfruna Street
Wolverhampton, WV1 1SB
U.K.
X.Zhang@wlv.ac.uk

Abstract: In Quantitative Risk Assessment, it is well known that the confidence of the probability result critically depends on the tail area of the random variables. The data in the tails are usually few in numbers in contrast with these in the centre, and it doesn't carry much weight in traditional analysis. Therefor, for getting a good fit of the distribution tail, a tail model based on Generalised Pareto Distribution (GPD) (Pickands, 1975) is proposed.

Meanwhile, as the failure area is normally very small, and in order to achieve accurate solution, an Asymptotic Important Sampling (AIS) is investigated in this paper. It is found to be an efficient technique for multivariate integration especially when the probability is small. The present paper is intended to introduce AIS to the tail model described above. An algorithm about the proposed AIS tail model in QRA is presented.

Keywords: Quantitative Risk Assessment, Asymptotic Important Sampling, Reliability Analysis.

1 Introduction

Recently, Quantitative Risk Assessment (QRA) is becoming more and more popular in reliability problem. But traditional analysis methods just concentrated on the central part of the variable distribution, most of it tries to obtain a good fit of the theoretical distribution in the central part of the observed data, rather than in the tail part.

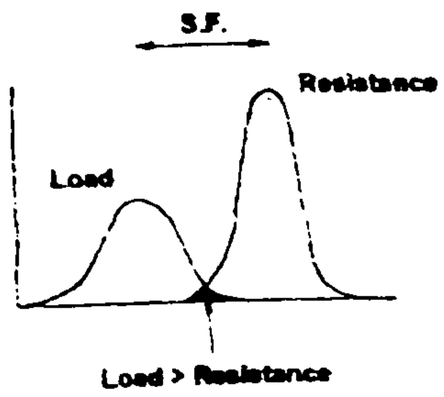


Fig.1 Failure area of QRA

As described in Fig.1, the area, which mostly effects the QRA result, lies in the distribution tail. Whether we can get an accurate result or not critically depends on if we can get a good fit of the tail area of the variable distributions. That is to say, the risk is often critically dependent on the upper and/or lower tail behaviour of one or a few basic uncertainties. If, for instance, the tail model of one these variables is altered slightly, it may very well be that the risk level associated with the model changes by an order of magnitude, even though the uncertainty modelling itself may be satisfactory from a statistical point of view. This situation is in sharp contrast with the advertised high level of accuracy now associated with (commercial) analysis tools available to solve the risk analysis given probabilistic assumptions regarding its basic uncertainties. This situation is, of course, highly undesirable. Professionals involved in quantitative risk assessment (QRA) are not well served by it, because it undermines their

claim to "correctness" or usability of their risk calculations.

As mentioned in reference [1], the aim of determining certain tail characteristics of random variables in the reliability-based design includes:

(1) the computation of high, or low, quantities of a random variable;
(2) the determination of very small probabilities of exceedance associated with fixed levels; or
(3) The accurate description of the tail of a probability distribution.

The latter objective is critically important in structural reliability analysis. If one fails to model the tail behaviour of the basic variables correctly, then the resulting reliability level is questionable.

The basic idea to achieve this aim is by weighting the approximation error with weights, which are larger in the distribution tail; it is ensured that the fit of the approximate distribution is good in the tail region. The selection of the weights is based on a risk criterion appropriate for structural reliability analysis.

Meanwhile, it is well known that the failure area is normally very small. How to achieve an accurate solution in this small area is an important objection. According to Reference [4], Asymptotic Important Sampling (AIS) technology is an efficient technique for multivariate integration especially when the probability is small. In QRA, almost all PDFs have these characteristics. Therefore, it would be efficient to use it into the tail model discussed above.

The key idea of AIS is based on the careful selection of a sampling density for subsequent use in an importance sampling scheme [5]. The selection is based on theoretical consideration of the structure of the integration domain in the original variable space. Due to the asymptotic properties of the sampling densities, the AIS technique becomes increasingly efficient (in terms of obtaining a given level of accuracy) as the failure probability becomes smaller.[4]

2 Probability Model

The safety of reliability problem can be expressed in terms of a safety margin equation as follows:

$$M = g(X_1, X_2, \cdots, X_n) \qquad (1)$$

Where M is the random safety margin and X_i are n uncertain variables (random variables), which may obey any form of probability distribution and which together govern whether or not failure occurs.

With model-based risk or reliability analysis, we refer to a framework where risk is assessed by means of small probabilities associated with a random variable Z which arises as the response of a model M. The model can be deterministic or stochastic. It is a common characteristic of such model-based analyses that the tail area of the response variable Z *is of interest, i.e. either the* upper range or the lower range of this variable.

In what follows, It is assumed that risk is quantifiable through the probability of exceedance or non-exceedance of a set value by the variable Z. This small probability or risk R_z can consequently be computed on the basis of the known model M and known stochastic properties of the uncertainties X_i.

The principal interest now is to examine the change of R_z as a function of the assumed models for the tails of the basic uncertainty variables X_i.

3 Distribution Tail Model

3.1 Tail heaviness index (THI)

It is important in the field of QRA to check whether an assumed distribution of an input variable X is indeed a fair and risk-consistent representation of reality. There is then a major difference between "central models" for which the use of classical statistical inference tools is appropriate, and "tail models" which are applicable to risk and reliability problems where the interest lies in the upper or lower tail area (see Fig.2). (See [2,3,12])

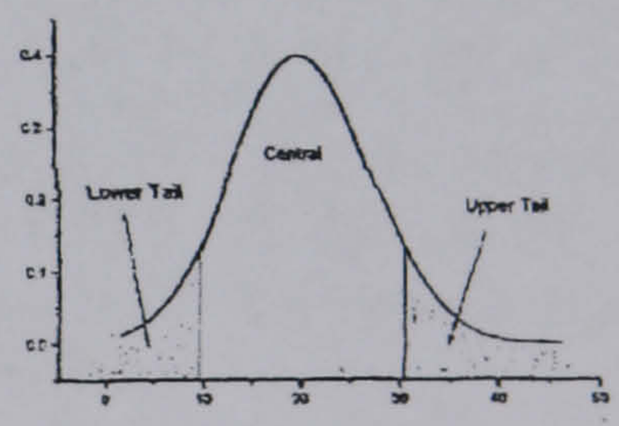


Fig.2 Lower/Upper Tail area

A useful parameter describing tail behaviour is the tail heaviness index (THI) (see, [5,10]) which basically expresses how heavy any portion of a probability density tail is with respect to the pure exponential tail (THI=0). It was introduced by Breiman et al. (see, [5]) and first used by Boos (see, [22]) in a comparative study of techniques

used to estimate large quantities. The idea is to benchmark heaviness against that of the exponential tail which is assigned a value of zero; it is negative for lighter than exponential tails (sub-exponential) and positive for heavier than exponential tails (super-exponential). The index is, generally, a function of the position on the tail, i.e. left-versus right-handed tail, as well as the exceedance probability q and the corresponding quantity

As described in reference [2], for a value x on the lower/upper tail of the random variable X, it is defined as

$$THI(x) = \frac{F(x)f'(x)}{f^2(x)} - 1 \qquad \text{Lower Tail}$$
$$THI(x) = \frac{-(1-F(x))f'(x)}{f^2(x)} - 1 \qquad \text{Upper Tail} \tag{2}$$

where F and f are the cumulative distribution function and the density of X, respectively. While for both types of tail the THI can most conveniently be related to the curvature of the minus-log-exceedance plot:

$$THI(x) = -\frac{L''(x)}{L'^2(x)} \tag{3}$$

where $L(x)$ is defined as:

$$L(x) = -\ln(F(x)) \qquad \textit{Lower tail}$$
$$L(x) = -\ln(1-F(x)) \qquad \textit{Upper tail} \tag{4}$$

3.2 Implementation

The generalised Pareto Distribution (GPD) (see, [2]) is a two-parameter distribution that contains the uniform, exponential, and Pareto distribution as special case. It has been used in wide variety of socio-economic applications and in failure problems of reliability studies.

Several methods of fitting the GPD have been proposed, such as the method of maximum likelihood (ML), the classical method of moments (MM) and the method of probability weighted moments (PWM) (see, [19,21). These methods have been compared in a simulation study by Hosking and Wallis (see [10]), where it was shown that unless the sample size is 500 or more, estimators derived by the MM or PWM were more reliable than those obtained by the ML method. One disadvantage of the classical MM is that the order of the moments that it uses to estimate the parameters of a given distribution is somewhat arbitrary. It is well known, for example, that the classical use of the mean and variance, to fit a two-parameter distribution, is theoretically better defendable for certain types of distributions than for others.

In reference [12], Pickands showed that GPD is very useful to capture the asymptotic behaviour of the tail. Therefore, GPD is now used to model these tail portions (see [2, 11]). For the ascending and descending parts of a distribution below the value u, the GPD distribution can be respectively written as follows:

$$F_{GPD}(x) = \left[1 + \xi\frac{(u-x)}{\sigma}\right]^{-1/\xi} \qquad (x < u) \qquad \text{Lower Tail}$$
$$F_{GPD}(x) = 1 - \left[1 + \xi\frac{(u-x)}{\sigma}\right]^{-1/\xi} \qquad (x > u) \qquad \text{Upper Tail} \tag{5}$$

Where u is the threshold value, σ is a positive parameter and ξ is, as it turns out [2], exactly equal to the THI for both upper and lower tails.

The optimal estimation of the GPD parameters can be achieved in the following steps:

- Estimate u on the basis of a mean residual life (MRL) diagram, as discussed in [2], by identifying the stationary linear trend in the tail.
- Minimise the sum of weighted square error(SWSE):

$$\text{SWSE} = \sum_{i \in Tail} (L(x_i) - L_{GPD}(x_i \mid u, \xi, \sigma))^2 \tag{6}$$

With respect to ξ and σ, where L is give by Eq.(4).

Obviously, the minimisation is consistent with the objective to obtain a risk estimate R_z that is weighted correctly for tail effects.

It is of course very important to perform a complete uncertainty analysis for this estimation procedure. Uncertainty is associated first with the model given in Eq(6) as well as with the parameters of the GPD themselves. Both can be treated in a way similar to that described in reference [3,11]. This allows a basic tail modelling uncertainty $\sigma^2_{L_i}(x)$ to be computed for the L_i value associated with each basic variable X_i.

3.3 Effect of Tails on QRA

The propagation of uncertainty through the model $g(X_1, X_2, \cdots, X_n)$ can be studied on a case by case basis. However, it can be shown [2] that in most practical cases, an error ΔL_i on the L function of the tail of the ith variable generates an error $\Delta(-\ln R_z)$ on the logarithm of the risk R_z approximately equal to:

$$\Delta(-\ln R_Z) \cong (1+THI_i)\Delta L_i(x_{PLM}) \qquad (5)$$

This approximation applies to asymptotic conditions, which are usually achieved when the risk R_Z is small and the number of variables is not excessive. The error on L_i is evaluated at the point of maximum likelihood (PML) associated with the model M. This is the point, which maximises the joint, log-likelihood of the variables $(X_1, X_2, \cdots X_n)$ subject to the constraint $g(X_1, X_2, \cdots X_n)$. This formula shows that the relative contribution of heavy tails (THI>0) is more pronounced that of light tails (THI<0), which includes, for instance, bounded tails.

The variance of the total uncertainty on $\ln R$, can now be determined as follows:

$$\sigma^2_{\ln R_Z} = \sum_{i=1}^{n}(1+THI_i)^2\sigma^2_{L_i} \qquad (6)$$

The question of how the model-based risk R_Z varies as a function of the tail heaviness index of the *j*th uncertainty variable X_i can also be addressed.

4 Asymptotic Importance Sampling (AIS)

4.1 The basic concept of Asymptotic Approximations for failure probability

Asymptotic approximations for Laplace-type integrals of the form

$$I(\beta) = \int_F e^{\beta^2 l(z)} dz \qquad (7)$$

where F is the failure domain: $F = \{z : g(z) \le 0\}$. In the case of failure probabilities, where $l(z)$ is scaled log-likelihood function of a probability density, there are some characteristics. Since in general the probability content of the failure domain is small and far away for the region where the PDF is large, the maximum of $l(z)$ is at some points of the boundary G of F. For such a maximum of function on a surface G, the Lagrange multiplier theorem gives some results about the structure of the Taylor expansion of l and g at the Point of maximum likelihood (PML) z^*. The gradients $\nabla l(z^*)$ and $\nabla g(z^*)$ are orthogonal to the tangential space of G at the PML and therefore

$$\nabla f(z^*) = \lambda \nabla g(z^*) \qquad (8)$$

Further, the first directional derivatives of f and g in the direction of the tangential space vanish. So, for a Taylor expansion of l at z^* in the region of F near this point, we need first the directional derivative expansion of l in the domain F, which is given by $-|\nabla l|$. Further we need the second directional derivatives into the direction of the tangential space of G at the PML of l and g to determine the behaviour of the function l on the surface G near z^*. With this information it's possible to derive the asymptotic approximation.

4.2 The basic multivariate asymptotic importance sampling density

A basic multivariate AIS density was given by Breitung (See [5]). The algorithm is described as follows.

By a simple and suitable translation and rotation we can introduce a new curvilinear co-ordinate system $x_1, x_2, \cdots, x_n$, such that:

(1) the PML lies at origin, that is, $x^* = (0, \cdots, 0)$;

(2) The unit vector $\hat{x}_1$ coincides with the direction of steepest descent at the PML:

$$\hat{x}_1 = -\left[\frac{g}{|\nabla g|}\right]_{z=z^*} \qquad (9)$$

(3) The remaining $n-1$ co-ordinates $x_2, \cdots, x_n$ are surface co-ordinates in the direction of the main curvatures on the limit state surface at PML.

(4) In practice, a third system of co-ordinates $y_1, y_2, \cdots, y_n$ needs to be introduced. This co-ordinate system is Cartesian with y_1 coinciding with x_1 and $y_2, \cdots, y_n$ being the projections of the $x_2, \cdots, x_n$ surface co-ordinate axes on the tangent hyper-plane at the PML.

By the above procedures, the basic problem is essentially similar to the univariate problem. An asymptotic approximation is to be found of the joint density $h_X(x)$ of the random vector X. Using key results from differential geometry and Laplace integrals, Breitung [2] developed the foundations for the following important result, for small P(F), approximately:

$$h_x(x) \sim \exp\left\{-|\nabla l|x_1 + \frac{1}{2}\sum_{i,j=2}^{n}\left(l_{ij} - \frac{|\nabla l|}{|\nabla g|}g_{ij}\right)xi\,x_j\right\}, \qquad x_1 > 0 \tag{10}$$

4.3 Importance Sampling

The basic random variables reliability problem consists of finding the probability that a set of n random variables Z, representing the physical and cognitive characteristics of a system, is contained within a failure domain $F = \{z : g(z) \le 0\}$

$$P(F) = \int_F f_z(z)dz \tag{11}$$

where $f_z(.)$ is the joint probability density function (PDF) of Z, and $g : \Re^n \to \Re$ is the state function of the system. The surface G={z: g(z)=0} that separates the safe set F^c from $\{z : g(z) < 0\}$ is referred to as the limit state surface (LSS).

In order to solve the integral (11) by statistical sampling, it is of utmost importance to control and to minimise the sampling error. The most frequently applied variance reduction scheme is that of importance sampling. To do so, a new set of random variables X is introduced with PDF h_X. The functional relationship between z and x may, in general, be non-linear and not even-to-one. However, in what follows, we will consider a transformation of points in failure domain with identical Jacobeans $J_{zx} = def[\partial z_i / \partial x_j]$ for each contributing disjoint set, such that

$$h_z(z(x)) = h_X(x)\left|J_{zx}\right|^{-1} \tag{12}$$

Consequently, the multivariate integral (11) can be written as:

$$P(F) = \int I(z \in F)\frac{f_z(z)}{h_z(z)}h_z(z)dz = \int I(z \in F)\frac{f_z(x)}{h_X(x)}|J_{zx}(x)|h_X(x)dx \tag{13}$$

with I denoting an indicator function equal to 1 for true arguments and zero otherwise. In other words, P(F) is determined as the expectation of a function w(x) with respect to $h_X(x)$:

$$P(F) = E_h[w(x)] \tag{14}$$

where

$$w(x) = \begin{cases} |J_{zx}(x)|f_z(x)/h_X(x) & if\ x \in F \\ 0 & elsewhere \end{cases} \tag{15}$$

The importance sampling PDF $h_X(x)$ must be nonzero over the failure domain, to ensure that the average $\overline{P}_m$ of randomly simulated values of the function $w(x_i)$ $(i = 1, \cdots, m)$ serves as an unbiased estimator of the failure probability:

$$\overline{P}_m = \frac{1}{m}\sum_{i=1}^{m} w(x_i) \tag{16}$$

with a mean and a variance equal to

$$E(\overline{P}_m) = P(F) \tag{17}$$

$$\mathrm{var}(\overline{P}_m) = \frac{1}{m}\mathrm{var}(w(x)) = \frac{1}{m}\int_{\Re^n}[w(x) - p(F)]^2 h_X(x)dx \tag{18}$$

where, minimum variance is achieved if w is as constant as possible over the failure domain and zero elsewhere. In practice, there are two key objectives: the first one is to construct a density h_X in such a way that it closely mimics the behaviour of f_z within F, and the second one is to ensure that of h_X is a PDF from which random variable vectors can easily be generated.

5 Implementation of AIS on Tail Model

In the original variable space, a simple linear transformation links z to the local Cartesian co-ordinates y at the PML. The transformation between y and x, however, requires additional attention. The practical implementation of (10) rest on the following processes.

The case of one limit state function g and a unique PML, z^* is considered first. Together with $g(z^*) = 0$ and $\nabla g \neq 0$, it is also assumed that the maximum is regular with respect to F, i.e. $\det C \neq 0$, where the $n-1$ by $n-1$ matrix C is equal to

$$C = -\left\{l_{ij} - \frac{|\nabla l|}{|\nabla g|}g_{ij}\right\}_{i,j=2,\cdots,n} \tag{13}$$

and where double subscripts denote second order derivatives.

At the PML(point of the maximum likelihood), y=0, the importance sampling density (10) is determined, the normalising constant being $|\nabla l|(2\pi)^{(n-1)/2}(\det C)^{1/2}$. A random sample $\{x_j\}$ (j=1,...,m) is generated, the first component being exponential with mean $|\nabla l|^{-1}$ and the (n-1) remaining components normal with mean 0 and covariance matrix C^{-1}. Let us concentrate on a random point S thus obtained in failure domain. First suppose that the limit state surface coincides exactly with its second-order approximation, such

that the co-ordinated $x_2, x_3, \cdots, x_n$ represent arc lengths on the paraboloid; they are the surface co-ordinates of the projection of A along the direction (9) onto the limit state surface. It follows that, in this case:

$$y_1 = -\frac{1}{2}\sum_{i=2}^{n} k_{gi} y_i^2$$
$$y_i = \frac{1}{k_{gi}}\eta^{-1}(k_{gi} x_i), \quad i = 2, \cdots, n \tag{20}$$

where k_{gi} are the principal curvatures in the directions $i(i = 1,2,\cdots,n)$. The function $a = \eta^{-1}(b)$ represents a numerically suitable inverse function of the non-dimensional arc length function $b = \eta(a)$ for a parabola:

$$\eta(a) = \frac{a}{2}\sqrt{1+a^2} + \frac{1}{2}\ln(a + \sqrt{1+a^2}), \qquad a \in \Re \tag{21}$$

Substituting the derivatives of the arc length function (21) and considering the additional fact that the Jacobean of the orthogonal transformation z->y is equal to one, it can be verified that:

$$|J_{zx}(x)| = |J_{zy}(x) J_{yx}(x)| = \prod_{i=2}^{n} [1 + (k_{gi} y_i(x_i))^2]^{-1/2} \tag{22}$$

where $y_i(x_i)$ is given by the Eq(14). It is now a sample matter to determine the value of the sample statistics $[w(x)]_j$ associated with each vector x_j (j=1,...,m) using Eq(15), (10) and (22).

6 Examples

Example 1: This example is for illustrating the accuracy and parameter sensitivity of using AIS, in the simple case of two independent log-normal variables z_1 and z_2, with parameters μ and σ, as shown in Table 1. The state function is as follows:

$$g(z_1, z_2) = \gamma^2 - z_1 z_2$$

the failure probability, the sensitivity of P(F) with respect to μ and σ, and the sensitivity of P(F) with respect to γ are determined respectively in this example.

Some of the parameters we need during the computation is as follows:

$$l(z_1, z_2) = c - 2\ln\sigma - \ln z_1 - \ln z_2 - \sum_{i=1}^{2} \frac{(\ln z_i - \mu)^2}{2\sigma^2}$$

$$PML = (\gamma, \gamma)^T,$$

$$\left|\frac{\nabla l}{\nabla g}\right| = c\gamma^{-2}, \ \text{where } c = \left[1 + \frac{\ln\gamma - \mu}{\sigma}\right]$$

$$\hat{x}_1 = \begin{Bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{Bmatrix} \quad k_{12} = \frac{1}{\sqrt{2}\gamma c\sigma^2}(1 - c\sigma^2),$$

$$k_{g2} = -\frac{1}{\sqrt{2}\gamma}$$

γ=4.15148, μ=1, σ=0.150

For this example, we can get the exact results by reducing as follows,

$$P(F) = \Phi\left(-\sqrt{2}\frac{\ln\gamma - \mu}{\sigma}\right),$$

$$\frac{\partial P(F)}{\partial \mu} = \frac{1}{\sqrt{\pi}\sigma} e^{-\left(\frac{\ln\gamma - \mu}{\sigma}\right)^2}$$

$$\frac{\partial P(F)}{\partial \sigma} = \frac{1}{\sqrt{\pi}\sigma^2}(\ln\gamma - \mu) e^{-\left(\frac{\ln\gamma - \mu}{\sigma}\right)^2}$$

$$\frac{\partial P(F)}{\partial \gamma} = \frac{1}{\sqrt{\pi}\sigma\gamma} e^{-\left(\frac{\ln\gamma - \mu}{\sigma}\right)^2}$$

Table 1 shows the comparison of the exact result and the one achieved by AIS, MCS and FORM/SORM, respectively. Using small quantity of samples, AIS method can achieve a good result. The parameter sensitivity is also analysed in this example, it shows that the probability is direct proportional to μ and σ, but inverse to γ. This example shows that AIS is more accurate than other methods.

Example 2: To investigate the characteristic of different methods under different variable space, the following example was employed. It provides a comparison with traditional methods Monte Carlo Simulation (MCS), First/Second Order Reliability Method (FORM/SORM).

The state function:

$$g(z) = \beta - z_1 + \frac{a}{n}\sum_{i=2}^{n}(\cosh z_i - 1)$$

$z_i (i = 1, \cdots, n)$ are n independent standard normally distributed random variables

We have:

$$l(z) = -\frac{1}{2}\sum_{i=1}^{n} z_i^2, \quad PML = (\beta,0,\cdots,0)^T, \quad \hat{x}_1 = (1,0,\cdots,0)^T$$

$$\frac{|\nabla l|}{|\nabla g|} = \beta, \qquad k_{li} = \frac{1}{\beta},\ k_{gi} = -\frac{a}{n},\ i = 2,\cdots,n$$

as an example, let $\beta = 3.5$, $a = 1$, the simulation results are shown in Table 2.

The error analysis is based on supposing MCS result (m=100,000) as accurate. This example shows that the AIS requires very few samples to obtain enough accuracy. The error increases as the dimensionality becomes higher. FORM can only be used in the pre-estimation, SORM is only acceptable for low dimensionality. In AIS, this problem can be solved by enlarging the important sampling.

7. Conclusion

Based on the geometry of the integration domain in the original variable space, an Importance sampling method is used to the tail model in this paper. Due to the asymptotic properties of the sampling densities, the AIS technique becomes increasingly efficient as the failure probability becomes smaller. The quality of the method is illustrated using several examples.

This method is well suited to improve reliability estimates obtained using FORM/SORM or MCS. It emphasises that the problem of determining small failure probabilities is related to extreme value analysis. Minimisation of sampling error can always be achieved in an asymptotic sense.

Acknowledgement

The authors would like to thank BG technology for the financial support of this research work.

References

[1]. H.O.Madsen, S. Krenk, N.C. Lind, *Methods of Structural Safety*, 1986, Prentice Hall.

[2]. Maes, M.A., *Tail Heaviness in Structural Reliability*, Proc. CERRA-ICASP, pp997-1002, 1995, Paris

[3]. Maes, M.A. and Breitung, K., 1993, *Reliability-Based Tail Estimation*, Proc. IUTAM Symposium on Probabilistic Mechanics, pp335-346, San Antonio, Texas, USA, 1993

[4]. Marc A. Maes, *Asymptotic Importance Sampling*, Structural Safety, Vol.12,1994.

[5]. K. Breitung, *Probability Approximations by Log-likelihood Maximisation*, J. Eng. Mech., 117(3) 1997.

[6]. SESAM Theory Manual of PROBAN, *General Purpose probability analysis program*

[7]. Maes, K.C. Gulati, D.L.Mckenna, P.R.Brand, D.B.Lewis, *Reliability Based Casing Design*, Journal of Energy Resources Technology, 1993

[8]. Averill M.Law, W.David Kelton, *Simulation Modelling and Analysis (Second Edition)*, McGraw-hill international Editions, 1991

[9]. Breiman, L., Stone, C.J., and Gins, J.D., 1979, *New Methods for Estimating Tail Probabilities and Extreme Value Distributions*, TSCorp, PD-A226-1, Santa Monica, California.

[10]. Hosking, J.R.M. and Wallis, J.R., 1987, *Parameter and quantile estimation for the generalised Pareto distribution*, Technometrics, 29(3): 339-349.

[11]. Castillo, E., 1988, *Extreme Value Theory in Engineering*, Academic Press, San Diego, California.

[12]. Pickands, J., *Statistical Interference Using Extreme Order Statistics*, Annals of Statistics, Vol. 3, pp.119-113, 1975

[13]. Ashkar, F., Ouarda, T.B.M.J, *On some methods of fitting the Generalised Pareto distribution*, J. Hydrology 177(1996), pp117-141

[14]. Ashkar, F., Ouarda, T.B.M.J., *The generalised method of moments for filling the generalised Pareto distribution*, Scientific Report STAT-14, Department of Mathematics, University of Moncton, N.H.,1994

[15]. Yolanda Carson, Anu Maria, *Simulation Optimisation: Methods and Applications*, 1997 winter Simulation Conference.

[16]. Van Montfort, M.A.J. and Witter, J.V., *The generalised Pareto distribution applied to rainfall depths*. Hydrol. Sci. J., 1986, 31(2): 151-162.

[17]. Tamano, T.,el al. A *new empirical formula for collapse resistance of commercial casing*, J. Energy Resources Technology, ASME, 1983.

[18]. *Bulletin on formulas and calculations for casing, tubing, drill pipe and line pipe properties*, API Bulletin 5C3, 6th edition, Oct 1994.

[19]. Greenwood, J.A., Landwehr, J.M., Matalas, N.C. and Wallis, J.R., *Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form*. Water Resour. Res., 1979, 15(5): 1049-1054.

[20]. Kendall, M., and Stuart, A., *The Advanced Theory of Statistics*, Vol. II, Fourth Edition, London, 1979

[21]. Hosking, J.R.M., *L-Moments: analysis and estimation of distributions using linear combination of order statistics*. R. Stat. Soc., Ser. B, 1990, 52(1): 105-124.

[22]. Boos, D.D., *Using Extreme Value Theory to Estimate Large Percentiles*, Technometrics, 1984, Vol 2, No. 1, pp. 33-39.

[23]. Kotz, S. and Johnson, N.L., *Pareto distribution. In: Encyclopedia of Statistical Sciences*, 1985, Vol. 6. Wiley, New York, pp. 568-574.

[24]. Smith, J.A., *Estimating the upper tail of flood frequency distributions*. Water Resour. Res., 1987, 23(8): 1657-1666.

[25]. Van Montfort, M.A.J. and Witter, J.V., *The generalised Pareto distribution applied to rainfall depths*. Hydrol. Sci. J., 1986, 31(2): 151-162.

[26]. Issa, J. A et al, *An Improved Design Equation For Tubular Collapse*, SPE 26317,Proceedings of SPE Annual Technical Conference, Houseton, OCT.,1993

[27]. Tokimasa, K. et al, *FEM Analysis of The Collapse Strength of A Tube*, J. of Pressure Vessel Technology, ASME, May, 1986

Table 1: comparison with exact results

		$P(F)$	Error	$\partial P/\partial\mu$	Error	$\partial P/\partial\sigma$	Error	$\partial P/\partial\gamma$	Error
Exact result		3.167E-5	—	1.261E-3	—	3.569E-3	—	-3.304E-4	—
MCS(m=10,000)		3.182E-5	0.47%	1.281E-3	1.54%	3.625E-3	1.57%	-3.547E-4	7.3%
FORM		3.428E-5	8.2%	1.420E-3	12.6%	3.207E-3	10.1%	-2.953E-4	10.6%
SORM		3.250E-5	2.62%	1.316E-3	4.36%	3.372E-3	2.72%	-3.012E-4	8.84%
AIS	m =50	3.176E-5	0.28%	1.261E-3	0	3.522E-3	1.30%	-3.018E-4	8.6%
	m =500	3.189E-5	0.29%	1.268E-3	0.50%	3.614E-3	1.26%	-3.024E-4	8.4%
	m =5000	3.166E-5	0.03%	1.262E-3	0.08%	3.566E-3	0.1%	-3.021E-4	8.5%

Table 2: comparison of different methods

	n =2	Error	n =10	Error	n =50	Error
MCS(m=100,000)	1.332E-4	—	4.229E-5	—	2.913E-5	—
FORM	2.326E-4	74%	2.326E-4	400%	2.326E-4	200%
SORM	1.403E-4	5.3%	6.028E-5	42%	4.434E-5	52%
AIS m =50	1.350E-4	1.4%	4.050E-5	4.2%	2.273E-5	22%
m =500	1.324E-4	0.6%	4.119E-5	2.6%	2.290E-5	21%
m =5000	1.325E-4	0.5%	4.163E-5	1.6%	2.293E-5	21%